

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:38:32 ; Search time 59 Seconds
(without alignments)
2078.400 Million cell updates/sec

Title: US-09-985-689A-1
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSYGLY.....EVOQVNPVGPOTFSLAIVN 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2247	100.0	434	5	AAW50080 Bacillus
2	2247	100.0	640	2	AAAY17090 Bacillus
3	2242	99.8	640	2	AAAY17091 Bacillus
4	2191	97.5	434	5	AAW50081 Bacillus
5	2183	97.2	639	2	AAAY17089 Bacillus
6	2155	95.9	639	2	AAAY17087 Bacillus
7	2155	95.9	640	2	AAAY17088 Bacillus
8	2143	95.4	434	5	AAW50085 Bacillus
9	2125.5	94.6	433	5	AAW50086 Bacillus
10	2125.5	94.6	641	2	AAW89547 Bacillus
11	2071	93.2	434	5	AAW50090 Bacillus
12	1998.5	88.9	433	5	AAW50084 Bacillus
13	1994.5	88.8	433	5	AAW50082 Bacillus
14	1989.5	88.5	636	2	AAW89548 Bacillus
15	1987.5	88.5	433	5	AAW50083 Bacillus
16	1986.5	88.4	433	2	AAW61495 Modified
17	1986.5	88.4	433	2	AAW61495 Modified
18	1986.5	88.4	433	2	AAW95698 Bacillus
19	1986.5	88.4	433	3	AAW63207 Amino aci
20	1986.5	88.4	433	3	AAW44619 Bacillus
21	1581.5	70.4	345	2	AAW62230 Subtilase
22	1581.5	70.4	345	2	AAW62230 Subtilase
23	452.5	20.1	659	2	AAW24121 Thermococ
24	452.5	20.1	659	2	AAW94840 WO9856926
25	414	18.4	412	2	AAW94836 Hyperther

26	414	18.4	522	2	AAW24122	Pyrococcus
27	414	18.4	522	2	AAW94838	Hyperther
28	414	18.4	654	2	AAW24129	Pyrococcus
29	414	18.4	654	2	AAW94841	Hyperther
30	401	17.8	659	2	AAW24123	Protease.
31	372	16.6	545	4	ABB09483	T. Yonsei
32	366	16.3	1079	4	ABB81180	Transglut
33	366	16.3	1079	6	ABU07391	Foreign P
34	346	15.4	520	2	AAW13666	Streptomy
35	346	15.4	734	2	AAW13667	Streptomy
36	346	15.4	823	2	AAW13668	DhpA-mel
37	338	15.0	1237	6	ABU11343	Protein e
38	310.5	13.8	806	2	AAW27481	RP-III re
39	306.5	13.6	519	6	ABP76735	Streptomy
40	306.5	13.6	19938	6	ABP76678	Streptomy
41	304.5	13.6	903	2	AAW87007	Streptomy
42	304.5	13.6	1398	2	AAW87008	Streptomy
43	304.5	13.6	1398	2	AAW24124	Protease.
44	304.5	13.6	1398	2	AAW94839	Pyrococcus
45	301.5	13.4	699	2	AAW08471	P. balust

ALIGNMENTS

RESULT 1
ID AAW50080 standard; protein; 434 AA.
XX
AC AAW50080;
XX
DT 12-AUG-2002 (first entry)
XX
DE Bacillus sp KSM-KP43 alkaline protease protein fragment.
XX
KW Alkaline protease; detergent; laundry; bleaching; dishwasher.
XX
OS Bacillus sp.
XX
PN EP1209233-A2.
XX
PD 29-MAY-2002.
XX
PF 22-NOV-2001; 2001EP-00127851.
XX
PR 22-NOV-2000; 2000JP-00355166.
PR 12-APR-2001; 2001JP-00114048.
XX
PA (KAOS) KAO CORP.
XX
PI Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
PI Okuda M, Saeki K;
XX
DR WPI; 2002-437518/47.
XX
PT New modified alkaline proteases useful in detergent compositions.
XX
PS Claim 1; Page 10-11; 25pp; English.
XX
CC This invention describes novel Bacillus sp. alkaline proteases useful in
CC detergent compositions, especially in laundry, bleaching or automatic
CC dishwasher detergents, the novel proteases have an increased detergency *
CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
CC sequence represents a fragment of the alkaline protease KP43 from
CC Bacillus sp strain KSM-KP43 which is used to create the modified protease
CC represented in AAW50090
XX
SQ Sequence 434 AA;

Query Match 100.0%; Score 2247; DB 5; Length 434;
Best Local Similarity 100.0%; Pred. No. 3.9e-159;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
 DB 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
 QY 61 NANDTNGHGHVAGSVLNGSTNKGAPQANLVFOSIMDSGGGLGGLPSNLQTLFQAYS 120
 DB 61 NANDTNGHGHVAGSVLNGSTNKGAPQANLVFOSIMDSGGGLGGLPSNLQTLFQAYS 120
 QY 121 AGARIHTNSWGAANVAGYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
 DB 121 AGARIHTNSWGAANVAGYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
 QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAQPFILSARSSLPDSSF 240
 DB 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAQPFILSARSSLPDSSF 240
 QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 DB 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 QY 301 PNGNQGWGRVTLDKSLNVAYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
 DB 301 PNGNQGWGRVTLDKSLNVAYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
 QY 361 SVTLVNDLVLITAPNGTQVGNDFTSYNDNDGRNNVNFVINAQSGTYTIEVOAYN 420
 DB 361 SVTLVNDLVLITAPNGTQVGNDFTSYNDNDGRNNVNFVINAQSGTYTIEVOAYN 420
 QY 421 VPVGPQTFSLAIVN 434
 DB 421 VPVGPQTFSLAIVN 434

RESULT 2

AAV17090
 ID AAV17090 standard; protein; 640 AA.

XX AAV17090;
 AC
 XX
 DT 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 DE Bacillus alkaline protease.
 XX
 KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent.
 XX
 OS Bacillus sp.
 XX
 PN WO9918218-A1.
 XX
 PD 15-APR-1999.
 XX
 PF 07-OCT-1998; 98WO-JP004528.
 XX
 PR 07-OCT-1997; 97JP-00274570.
 XX
 PA (KAOS) KAO CORP.
 XX
 PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX
 DR WPI; 1999-287736/27.
 DR N-PSDB; AAX37278.
 XX
 XX Alkali protease from Bacillus used in washing powders.
 PT
 XX
 PS Disclosure; Page 58-63; 71pp; Japanese.
 XX

CC The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it

CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease. (Updated
 CC on 20-MAR-2003 to correct DR field.)

XX
 SQ Sequence 640 AA;

Query Match 100.0%; Score 2247; DB 2; Length 640;
 Best Local Similarity 100.0%; Pred. No. 6.5e-159;
 Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
 DB 207 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 266
 QY 61 NANDTNGHGHVAGSVLNGSTNKGAPQANLVFOSIMDSGGGLGGLPSNLQTLFQAYS 120
 DB 267 NANDTNGHGHVAGSVLNGSTNKGAPQANLVFOSIMDSGGGLGGLPSNLQTLFQAYS 326
 QY 121 AGARIHTNSWGAANVAGYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
 DB 327 AGARIHTNSWGAANVAGYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 386
 QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAQPFILSARSSLPDSSF 240
 DB 387 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAQPFILSARSSLPDSSF 446
 QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 DB 447 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 506
 QY 301 PNGNQGWGRVTLDKSLNVAYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
 DB 507 PNGNQGWGRVTLDKSLNVAYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 566
 QY 361 SVTLVNDLVLITAPNGTQVGNDFTSYNDNDGRNNVNFVINAQSGTYTIEVOAYN 420
 DB 567 SVTLVNDLVLITAPNGTQVGNDFTSYNDNDGRNNVNFVINAQSGTYTIEVOAYN 626
 QY 421 VPVGPQTFSLAIVN 434
 DB 627 VPVGPQTFSLAIVN 640

RESULT 3

AAV17091
 ID AAV17091 standard; protein; 640 AA.

XX AAV17091;
 AC
 XX
 DT 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 DE Bacillus alkaline protease.
 XX
 KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent.
 XX
 OS Bacillus sp.
 XX
 PN WO9918218-A1.
 XX
 PD 15-APR-1999.
 XX
 PF 07-OCT-1998; 98WO-JP004528.
 XX
 PR 07-OCT-1997; 97JP-00274570.

XX PA (KAOS) KAO CORP.
 XX PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 XX PI Shikata S, Nomura M;
 XX DR WPI; 1999-287736/27.
 XX DR N-PSDB; AAX37279.
 XX PT Alkali protease from Bacillus used in washing powders.
 XX PS Disclosure; Page 63-68; 7lpp; Japanese.
 XX CC The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease. (Updated
 CC on 20-MAR-2003 to correct DR field.)
 XX SQ Sequence 640 AA;
 Query Match 99.8%; Score 2242; DB 2; Length 640;
 Best Local Similarity 99.8%; Pred. No. 1.5e-158;
 Matches 433; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 NDVARGIVKADVAQSSYGLYGQGGQIVAVADTGLDTGRNDSMHEAFRGKITALYALGRTN 60
 DB 207 NDVARGIVKADVAQSSYGLYGQGGQIVAVADTGLDTGRNDSMHEAFRGKITALYALGRTN 266
 QY 61 NANDTNGHGHVAGSVLNGSTNKGMAQANLVFQSIIMDSGGGLGGLPSNLQTLFSQAYS 120
 DB 267 NANDTNGHGHVAGSVLNGSTNKGMAQANLVFQSIIMDSGGGLGGLPSNLQTLFSQAYS 326
 QY 121 AGARIHTNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 DB 327 AGARIHTNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 386
 QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVWAPGTFILSARSSLPDSSF 240
 DB 387 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVWAPGTFILSARSSLPDSSF 446
 QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLGY 300
 DB 447 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLGY 506
 QY 301 PNGNQGWGRVTLDKSLNVAAYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
 DB 507 PNGNQGWGRVTLDKSLNVAAYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 566
 QY 361 SVTLVNDLVLVITAPNGTQYVGNDFTPSYNDWGRNENVENFINAPQSGTYYTIEVQAYN 420
 DB 567 SVTLVNDLVLVITAPNGTQYVGNDFTPSYNDWGRNENVENFINAPQSGTYYTIEVQAYN 626
 QY 421 VPVGQPTFSLAIVN 434
 DB 627 VPVGQPTFSLAIVN 640
 RESULT 4
 AAM50081
 ID AAM50081 standard; protein; 434 AA.
 XX AC
 XX AC AAM50081;
 XX XX

DT 12-AUG-2002 (first entry)
 XX Bacillus sp-KSM-KP9860 alkaline protease protein fragment.
 XX KW Alkaline protease; detergent; laundry; bleaching; dishwasher.
 XX OS Bacillus sp.
 XX PN EP1209233-A2.
 XX PD 29-MAY-2002.
 XX PF 22-NOV-2001; 2001EP-00127851.
 XX PR 22-NOV-2000; 2000JP-00355166.
 XX PR 12-APR-2001; 2001JP-00114048.
 XX PA (KAOS) KAO CORP.
 XX PI Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
 XX PI Okuda M, Saeki K;
 XX DR WPI; 2002-437518/47.
 XX PT New modified alkaline proteases useful in detergent compositions.
 XX PS Claim 5; Page 12-13; 25pp; English.
 XX CC This invention describes novel Bacillus sp. alkaline proteases useful in
 CC detergent compositions, especially in laundry, bleaching or automatic
 CC dishwasher detergents. The novel proteases have an increased detergency %
 CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
 CC sequence represents a fragment of the alkaline protease KP9860 from
 CC Bacillus sp strain KSM-KP9860 described in the method of the invention
 XX SQ Sequence 434 AA;
 Query Match 97.5%; Score 2191; DB 5; Length 434;
 Best Local Similarity 96.5%; Pred. No. 5.8e-155;
 Matches 419; Conservative 13; Mismatches 2; Indels 0; Gaps 0;
 QY 1 NDVARGIVKADVAQSSYGLYGQGGQIVAVADTGLDTGRNDSMHEAFRGKITALYALGRTN 60
 DB 1 NDVARGIVKADVAQSSYGLYGQGGQIVAVADTGLDTGRNDSMHEAFRGKITALYALGRTN 60
 QY 61 NANDTNGHGHVAGSVLNGSTNKGMAQANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS 120
 DB 61 NANDTNGHGHVAGSVLNGSTNKGMAQANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS 120
 QY 121 AGARIHTNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 DB 121 AGARIHTNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVWAPGTFILSARSSLPDSSF 240
 DB 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVWAPGTFILSARSSLPDSSF 240
 QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLGY 300
 DB 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLGY 300
 QY 301 PNGNQGWGRVTLDKSLNVAAYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
 DB 301 PNGNQGWGRVTLDKSLNVAAYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
 QY 361 SVTLVNDLVLVITAPNGTQYVGNDFTPSYNDWGRNENVENFINAPQSGTYYTIEVQAYN 420
 DB 361 SVTLVNDLVLVITAPNGTQYVGNDFTPSYNDWGRNENVENFINAPQSGTYYTIEVQAYN 420
 QY 421 VPVGQPTFSLAIVN 434
 DB 421 VPVGQPTFSLAIVN 434

42x

RESULT 5
 AAY17089
 ID AAY17089 standard; protein; 639 AA.
 XX
 AC AAY17089;
 XX
 DT 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 XX Bacillus alkaline protease.
 DE Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent.
 XX
 OS Bacillus sp.
 XX
 XX WO9918218-A1.
 XX
 PD 15-APR-1999.
 XX
 PF 07-OCT-1998; 98WO-JP004528.
 XX
 PR 07-OCT-1997; 97JP-00274570.
 XX
 PA (KAOS) KAO CORP.
 XX
 PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX
 DR WPI; 1999-287736/27.
 DR N-PSDB; AAX37277.
 XX
 PT Alkali protease from Bacillus used in washing powders.
 XX
 PS Disclosure; Page 53-58; 71pp; Japanese.
 XX
 CC The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease. (Updated
 CC on 20-MAR-2003 to correct DR field.)
 XX
 SQ Sequence 639 AA;
 Query Match 97.2%; Score 2183; DB 2; Length 639;
 Best Local Similarity 96.3%; Pred. No. 3.9e-154;
 Matches 418; Conservative 13; Mismatches 3; Indels 0; Gaps 0;
 QY 1 NDVARGIVKADVAQSSYGLYQCOQIVAVADTGLDGRNDSMHEAPRGKITALYALGRTN 60
 DB 206 NDVARGIVKADVAQSSYGLYQCOQIVAVADTGLDGRNDSMHEAPRGKITALYALGRTN 265
 QY 61 NANTDTHGHGTHVAGSVLNGSTNKGAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
 DB 266 NANTDTHGHGTHVAGSVLNGATNKGAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAFS 325
 QY 121 AGARIHTNSGAAVNGAYTTDSNRVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 DB 326 AGARIHTNSGAAVNGAYTTDSNRVDYVRKNDMTILFAAGNERPNGGTISAPGTAKNAI 385
 QY 181 TVGATENLRFSFGSYADNINHVAFSSRGPTKDGRIKPDVMAPGTGILSARSSLAPDSF 240
 DB 386 TVGATENLRFSFGSYADNINHVAFSSRGPTKDGRIKPDVMAPGTGILSARSSLAPDSF 445

241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSILLKALITAGADIGLY 300
 446 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSILLKALITAGADVGLGY 505
 301 PNGNQGWGRVTLDKSLNVAAYVNESSLSLTSQKATYSFTATAGKPLKISLVNSDAPASTTA 360
 506 PNGNQGWGRVTLDKSLNVAAYVNESSALSSTQKATYFTATAGKPLKISLVNSDAPASTTA 565
 361 SVTLVNDLDLIVITAPNGTQVYVNDFTSPYNDNDGRNVNENFINAPQSGTYTIEVQAYN 420
 566 SVTLVNDLDLIVITAPNGTQVYVNDFTSPYNDNDGRNVNENFINAPQSGTYTIEVQAYN 625
 421 VPVGPQTFSLAIVN 434
 526 VPVGPQNFSLAIVN 639
 427

RESULT 6
 AAY17087
 ID AAY17087 standard; protein; 639 AA.
 XX
 AC AAY17087;
 XX
 DT 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 XX An alkaline protease sequence from Bacillus species.
 DE Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent.
 XX
 OS Bacillus sp.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1..639
 FT /note= "all residues indicated as Xaa are arbitrary amino
 FT acids"
 XX
 WO9918218-A1.
 15-APR-1999.
 07-OCT-1998; 98WO-JP004528.
 07-OCT-1997; 97JP-00274570.
 (KAOS) KAO CORP.
 Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 Shikata S, Nomura M;
 WPI; 1999-287736/27.
 N-PSDB; AAX37277.
 Alkali protease from Bacillus used in washing powders.
 Disclosure; Page 53-58; 71pp; Japanese.
 The invention relates to alkaline proteases produced by strains of
 Bacillus. The proteases ability to digest casein is not inhibited by
 oleic acid and they have a high stability to oxidising agents. The
 alkaline protease of the invention has the following properties: (a) it
 is active over the pH range 4-13 and has at least 80% of its optimum
 activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
 stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
 its ability to digest casein is not inhibited by oleic acid; (e) it has
 molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 used as enzymes in washing compositions for use in automatic dishwashers
 and for washing clothes. The stability to oxidising agents allows the
 enzyme to be an effective component of washing compositions including
 bleaches. The present sequence represents an alkaline protease. (Updated
 on 20-MAR-2003 to correct DR field.)


```

XX SQ Sequence 639 AA;
Query Match 95.9%; Score 2155; DB 2; Length 639;
Best Local Similarity 96.3%; Pred. No. 4.7e-152;
Matches 418; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 1 NDVARGIVKADVAGSSYGLYGQGIIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60
Db 206 NDVARGIVKADVAGSSYGLYGQGIIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 265
QY 61 NANTDNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIIMDSGGGLGGLPSNLQTLFQAYS 120
Db 266 NANTDNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIIMDSGGGLGGLPSNLQTLFQAYS 325
QY 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTTISAPGTAKNAI 180
Db 326 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTTISAPGTAKNAI 385
QY 181 TVGATENLRPSFGSYADNINHVAFSSRGTKGRIPKDVNAPGTFTILSARSLAPDSSF 240
Db 386 TVGATENLRPSFGSYADNINHVAFSSRGTKGRIPKDVNAPGTFTILSARSLAPDSSF 445
QY 241 WANHDSKYAYMGTSMATPIVAGNVAOLREHFVKNRGITPKPSLLKALIAAGADIGLY 300
Db 446 WANHDSKYAYMGTSMATPIVAGNVAOLREHFVKNRGITPKPSLLKALIAAGADIGLY 505
QY 301 PNGNQGWGRVTLDKSLNVAVYVNESSLSLSTSQKATYSPTATAGPLKISLVWSDAPASTTA 360
Db 506 PNGNQGWGRVTLDKSLNVAVYVNESSLSLSTSQKATYSPTATAGPLKISLVWSDAPASTTA 565
QY 361 SVTLVNDLDLVITAPNGTQVYVGNDFTPSYNDNDGRNNVENVFINAPQSGTYTIEVQAYN 420
Db 566 SVTLVNDLDLVITAPNGTQVYVGNDFTPSYNDNDGRNNVENVFINAPQSGTYTIEVQAYN 625
QY 421 VPVGPQTFSLAIVN 434
Db 626 VPVGPQTFSLAIVN 639

RESULT 7
AAAY17088 standard; protein; 640 AA.
XX ID AAAY17088
XX AC AAAY17088;
XX DT 20-MAR-2003 (revised)
XX DT 21-JUL-1999 (first entry)
XX An alkaline protease sequence from Bacillus species.
XX DE Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
XX KW washing composition; oxidising agent.
XX OS Bacillus sp.
XX FH Key Location/Qualifiers
XX FT Misc-difference 1..640
XX FT /note= "all residues indicated as Xaa are arbitrary amino
XX FT acids"
XX WO9918218-A1.
XX PD 15-APR-1999.
XX PF 07-OCT-1998; 98WO-JP004528.
XX PR 07-OCT-1997; 97JP-00274570.
XX PA (KAOS ) KAO CORP.
XX PI Takaiwa M, Okuda M, Saeiki K, Kubota H, Hitomi J, Kageyama Y;
XX PI Shikata S, Nomura M;

```

```

XX WPI; 1999-287736/27.
XX N-PSDB; AAX37278.
XX Alkali protease from Bacillus used in washing powders.
XX Claim 3; Page 50-53; 71pp; Japanese.
XX The invention relates to alkaline proteases produced by strains of
XX Bacillus. The proteases ability to digest casein is not inhibited by
XX oleic acid and they have a high stability to oxidising agents. The
XX alkaline protease of the invention has the following properties: (a) it
XX is active over the pH range 4-13 and has at least 80% of its optimum
XX stable over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
XX its ability to digest casein is not inhibited by oleic acid; (e) it has
XX molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
XX used as enzymes in washing compositions for use in automatic dishwashers
XX and for washing clothes. The stability to oxidising agents allows the
XX enzyme to be an effective component of washing compositions including
XX bleaches. The present sequence represents an alkaline protease of the
XX invention. (Updated on 20-MAR-2003 to correct DR field.)
XX Sequence 640 AA;
Query Match 95.9%; Score 2155; DB 2; Length 640;
Best Local Similarity 96.3%; Pred. No. 4.7e-152;
Matches 418; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 1 NDVARGIVKADVAGSSYGLYGQGIIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60
Db 207 NDVARGIVKADVAGSSYGLYGQGIIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 266
QY 61 NANTDNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIIMDSGGGLGGLPSNLQTLFQAYS 120
Db 267 NANTDNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIIMDSGGGLGGLPSNLQTLFQAYS 326
QY 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTTISAPGTAKNAI 180
Db 327 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTTISAPGTAKNAI 386
QY 181 TVGATENLRPSFGSYADNINHVAFSSRGTKGRIPKDVNAPGTFTILSARSLAPDSSF 240
Db 387 TVGATENLRPSFGSYADNINHVAFSSRGTKGRIPKDVNAPGTFTILSARSLAPDSSF 446
QY 241 WANHDSKYAYMGTSMATPIVAGNVAOLREHFVKNRGITPKPSLLKALIAAGADIGLY 300
Db 447 WANHDSKYAYMGTSMATPIVAGNVAOLREHFVKNRGITPKPSLLKALIAAGADIGLY 506
QY 301 PNGNQGWGRVTLDKSLNVAVYVNESSLSLSTSQKATYSPTATAGPLKISLVWSDAPASTTA 360
Db 507 PNGNQGWGRVTLDKSLNVAVYVNESSLSLSTSQKATYSPTATAGPLKISLVWSDAPASTTA 566
QY 361 SVTLVNDLDLVITAPNGTQVYVGNDFTPSYNDNDGRNNVENVFINAPQSGTYTIEVQAYN 420
Db 567 SVTLVNDLDLVITAPNGTQVYVGNDFTPSYNDNDGRNNVENVFINAPQSGTYTIEVQAYN 626
QY 421 VPVGPQTFSLAIVN 434
Db 627 VPVGPQTFSLAIVN 640

RESULT 8
AAAM50085 standard; protein; 434 AA.
XX ID AAAM50085
XX AC AAAM50085;
XX DT 12-AUG-2002 (first entry)
XX DE Bacillus sp alkaline protease protein A-1 fragment.
XX KW Alkaline protease; detergent; laundry; bleaching; dishwasher.

```

```

XX OS Bacillus sp.
XX PN EPI209233-A2.
XX PD 29-MAY-2002.
XX XX
XX XX 22-NOV-2001; 2001EP-00127851.
XX PF 22-NOV-2001; 2000JP-00355166.
XX PR 12-APR-2001; 2001JP-00114048.
XX XX
XX PA (KAOS ) KAO CORP.
XX XX
XX PI Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
XX PI Okuda M, Saeki K;
XX XX
XX DR WPI; 2002-437518/47.
XX XX
XX PT New modified alkaline proteases useful in detergent compositions.
XX XX
XX PS Claim 5; Page 18-19; 25pp; English.
XX XX
XX CC This invention describes novel Bacillus sp. alkaline proteases useful in
XX CC detergent compositions, especially in laundry, bleaching or automatic
XX CC dishwasher detergents. The novel proteases have an increased detergency &
XX CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
XX CC sequence represents a fragment of the alkaline protease A-1 from Bacillus
XX CC sp NCIB2289 described in the method of the invention
XX XX
XX SQ Sequence 434 AA;
XX
Query Match 95.4%; Score 2143; DB 5; Length 434;
Best Local Similarity 93.5%; Pred. No. 2.2e-151;
Matches 406; Conservative 19; Mismatches 9; Indels 0; Gaps 0;
QY 1 NDVARGIVKADVAQSSYGLYGGQGIIVAVADTGLDTCGRNDSMHEAFRGKITAIYALGRIN 60
Db 1 NDVARGIVKADVAQSSYGLYGGQGIIVAVADTGLDTCGRNDSMHEAFRGKITAIYALGRIN 60
QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNLQTLFSQAYS 120
Db 61 NANDPNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNVSTLFSQAYS 120
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db 121 AGARIHTNSGAPVNGAYTTDSRNVDYVRKNDMAVLFAAGNEGPNNGGTISAPGTAKNAI 180
QY 181 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIPKDVMAFGTIFLSARSSLAPDSFS 240
Db 181 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIPKDVMAFGTIFLSARSSLAPDSFS 240
QY 241 WANHDSKYATMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLGY 300
Db 241 WANHDSKYATMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLGY 300
QY 301 PNGNQGWGRVTLDKSLNVAIVNBSLSSTSQATYSFTATAGKPLKISLVWSDAPASTTA 360
Db 301 PNGNQGWGRVTLDKSLNVAIVNBSLSSTSQATYSFTATAGKPLKISLVWSDAPASTTA 360
QY 361 SVTLVNDLDDLVITAPNGTQVGNDFTSYNDNDWGRNENVENFINAPQSGTITIEVQAYN 420
Db 361 SVTLVNDLDDLVITAPNGTQVGNDFTSYNDNDWGRNENVENFINAPQSGTITIEVQAYN 420
QY 421 VPVGQPTFSLAIYN 434
Db 421 VPQGPQAFSLAIYN 434

```

RESULT 9
AAM50086
ID AAM50086 standard; protein; 433 AA.
XX

```

AC AAM50086;
XX
XX DT 12-AUG-2002 (first entry)
XX DE Bacillus sp alkaline protease protein A-2 fragment.
XX KW Alkaline protease; detergent; laundry; bleaching; dishwasher.
XX OS Bacillus sp.
XX PN EPI209233-A2.
XX PD 29-MAY-2002.
XX XX
XX XX 22-NOV-2001; 2001EP-00127851.
XX PF 22-NOV-2001; 2000JP-00355166.
XX PR 12-APR-2001; 2001JP-00114048.
XX XX
XX PA (KAOS ) KAO CORP.
XX XX
XX PI Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
XX PI Okuda M, Saeki K;
XX XX
XX DR WPI; 2002-437518/47.
XX XX
XX PT New modified alkaline proteases useful in detergent compositions.
XX XX
XX PS Claim 5; Page 20-21; 25pp; English.
XX XX
XX CC This invention describes novel Bacillus sp. alkaline proteases useful in
XX CC detergent compositions, especially in laundry, bleaching or automatic
XX CC dishwasher detergents. The novel proteases have an increased detergency &
XX CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
XX CC sequence represents a fragment of the alkaline protease A-2 from Bacillus
XX CC sp NCIB12513 described in the method of the invention
XX XX
XX SQ Sequence 433 AA;
XX
Query Match 94.6%; Score 2125.5; DB 5; Length 433;
Best Local Similarity 93.5%; Pred. No. 4.4e-150;
Matches 406; Conservative 19; Mismatches 8; Indels 1; Gaps 1;
QY 1 NDVARGIVKADVAQSSYGLYGGQGIIVAVADTGLDTCGRNDSMHEAFRGKITAIYALGRIN 60
Db 1 NDVARGIVKADVAQSNFGLYGGQGIIVAVADTGLDTCGRNDSMHEAFRGKITAIYALGRIN 60
QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNLQTLFSQAYS 120
Db 61 NANDPNGHGHVAGSVLNG-ATNKGMAPOANLVFQSIMDSGGGLGSLPANLQTLFSQAYS 119
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db 120 AGARIHTNSGAPVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIPKDVMAFGTIFLSARSSLAPDSFS 240
Db 180 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIPKDVMAFGTIFLSARSSLAPDSFS 239
QY 241 WANHDSKYATMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLGY 300
Db 240 WANHDSKYATMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLGY 299
QY 301 PNGNQGWGRVTLDKSLNVAIVNBSLSSTSQATYSFTATAGKPLKISLVWSDAPASTTA 360
Db 300 PNGNQGWGRVTLDKSLNVAIVNBSLSSTSQATYSFTATAGKPLKISLVWSDAPASTTA 359
QY 361 SVTLVNDLDDLVITAPNGTQVGNDFTSYNDNDWGRNENVENFINAPQSGTITIEVQAYN 420
Db 360 SVTLVNDLDDLVITAPNGTQVGNDFTSYNDNDWGRNENVENFINAPQSGTITIEVQAYN 419
QY 421 VPVGQPTFSLAIYN 434
XX

```


CC detergent compositions, especially in laundry, bleaching or automatic
 CC dishwasher detergents. The novel proteases have an increased detergency &
 CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
 CC sequence represents a fragment of the alkaline protease SD-521 from
 CC Bacillus sp strain SD-521 (FERM BP-11162) described in the method of the
 CC invention
 XX
 XX
 SQ Sequence 433 AA;

Query Match 88.9%; Score 1998.5; DB 5; Length 433;
 Best Local Similarity 87.8%; Pred. No. 1.3e-140; Indels 1; Gaps 1;
 Matches 381; Conservative 28; Mismatches 24;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTGRNDSMHEAFRGKITALYALGRN 60
 DB 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDTGRNDSMHEAFRGKITALYALGRN 60

QY 61 NANTNGHGTHTVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNLQTLFSQAYS 120
 DB 61 NANTNGHGTHTVAGSVLNG-ALNKGMAPOANLVFQSIMDSGGGLGSLPSNLQTLFSQAWN 119

QY 121 AGARIHTNSGAAVNGAYTTDSRIWDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 DB 120 AGARIHTNSGAPVNGAYTANSQVDEVVRNNDMTVLFAAGNEGPNSTISAPGTAKNAI 179

QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKGRKIPDVMAPOTFILSARSLAPDSF 240
 DB 180 TVGATENYRPSFGSLADNPNHIAQFSSRGATRDGRKIPDVAPGTFTILSARSLAPDSF 239

QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGADIGLY 300
 DB 240 WANNYSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGADVGLGY 299

QY 301 PNGNQGWGRVTLDKSLNVAIVYNESSLSSTOKATYSFTATAGKPLKISLVMSDAPASTTA 360
 DB 300 PSGDQGWGRVTLDKSLNVAIVYNEATALTQKATYSFQAQAKPLKISLVWTDAPGSTTA 359

QY 361 SVTLVNDLDLVTAPNGTQVYVNDFTSPYNDWGRNENVENFINAPQSGTYTIEVQAYN 420
 DB 360 SYTLVNDLDLVTAPNGQKYVGNDFSYPDNNWGRNENVENFINAPQSGTYTIEVQAYN 419

QY 421 VPVGQPTFSLAIVN 434
 DB 420 VPSGPQRFSLAIVH 433

RESULT 13
 AAM50082
 ID AAM50082 standard; protein; 433 AA.
 AC AAM50082;
 XX
 XX
 DT 12-AUG-2002 (first entry)
 XX
 XX Bacillus sp D6- (FERM P1592) alkaline protease protein fragment.
 DE
 XX Alkaline protease; detergent; laundry; bleaching; dishwasher.
 KW
 XX Bacillus sp.
 OS
 XX
 XX EP1209233-A2.
 PN
 XX
 XX 29-MAY-2002.
 PD
 XX
 XX 22-NOV-2001; 2001EP-00127851.
 PF
 XX
 XX 22-NOV-2000; 2000JP-00355166.
 PR
 XX 12-APR-2001; 2001JP-00114048.
 XX
 XX
 XX (KAOS) KAO CORP.
 PA
 XX
 XX Harada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
 PI Okuda M, Saeki K;

XX WPI; 2002-437518/47.
 DR
 XX New modified alkaline proteases useful in detergent compositions.
 PT
 XX Claim 5; Page 13-15; 25pp; English.
 PS
 XX This invention describes novel Bacillus sp. alkaline proteases useful in
 CC detergent compositions, especially in laundry, bleaching or automatic
 CC dishwasher detergents. The novel proteases have an increased detergency &
 CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
 CC sequence represents a fragment of the alkaline protease E-1 from Bacillus
 CC sp strain D6- (FERM-P1592) described in the method of the invention
 XX
 XX Sequence 433 AA;

Query Match 88.8%; Score 1994.5; DB 5; Length 433;
 Best Local Similarity 87.6%; Pred. No. 2.6e-140; Indels 1; Gaps 1;
 Matches 380; Conservative 28; Mismatches 25;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTGRNDSMHEAFRGKITALYALGRN 60
 DB 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDTGRNDSMHEAFRGKITALYALGRN 60

QY 61 NANTNGHGTHTVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNLQTLFSQAYS 120
 DB 61 NANTNGHGTHTVAGSVLNG-ALNKGMAPOANLVFQSIMDSGGGLGSLPSNLQTLFSQAWN 119

QY 121 AGARIHTNSGAAVNGAYTTDSRIWDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 DB 120 AGARIHTNSGAPVNGAYTANSQVDEVVRNNDMTVLFAAGNEGPNSTISAPGTAKNAI 179

QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKGRKIPDVMAPOTFILSARSLAPDSF 240
 DB 180 TVGATENYRPSFGSLADNPNHIAQFSSRGATRDGRKIPDVAPGTFTILSARSLAPDSF 239

QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGADIGLY 300
 DB 240 WANNYSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGADVGLGY 299

QY 301 PNGNQGWGRVTLDKSLNVAIVYNESSLSSTOKATYSFTATAGKPLKISLVMSDAPASTTA 360
 DB 300 PSGDQGWGRVTLDKSLNVAIVYNEATALTQKATYSFQAQAKPLKISLVWTDAPGSTTA 359

QY 361 SVTLVNDLDLVTAPNGTQVYVNDFTSPYNDWGRNENVENFINAPQSGTYTIEVQAYN 420
 DB 360 SYTLVNDLDLVTAPNGQKYVGNDFSYPDNNWGRNENVENFINAPQSGTYTIEVQAYN 419

QY 421 VPVGQPTFSLAIVN 434
 DB 420 VPSGPQRFSLAIVH 433

RESULT 14
 AAM89548
 ID AAM89548 standard; protein; 636 AA.
 AC AAM89548;
 XX
 XX
 DT 12-APR-1999 (first entry)
 XX
 XX Bacillus sp. alkaline protease Y.
 DE
 XX Alkaline protease Y; detergent; surfactant; leather processing;
 KW debittering; flavour.
 KW
 XX Bacillus sp.
 CS
 XX WO9856927-A2.
 PN
 XX 17-DEC-1998.
 PD
 XX
 XX 09-JUN-1998; 98WO-US012005.
 PF

XX 12-JUN-1997; 97US-00873479.
 XX (NOVO) NOVO NORDISK BIOTECH INC.
 XX Sloma A, Christianson L;
 XX WPI; 1999-080908/07.
 XX Novel protease from *Bacillus subtilis* LC20 - useful in laundry and
 PT dishwashing detergents and for leather processing.
 XX Claim 3; Page 55-56; 77pp; English.
 XX This is the amino acid sequence of a *Bacillus* sp. alkaline protease Y
 CC that is said to have good alkali and surfactant resistance and improved
 CC detergency. It shows 77% identity to a newly isolated protease (see
 CC AA89547) of *Bacillus* sp. JPI70 (NCIB 12513). The invention provides
 CC vectors, recombinant host cells and methods for the recombinant
 CC production of such proteases. The protease are used in laundry and
 CC dishwashing detergents, for institutional and industrial cleaning, and
 CC for leather processing, as well as for debittering and enhancing the
 CC degree of hydrolysis of protein hydrolysates, for flavour development
 CC through hydrolysis of proteins, degradation of undesired peptides and in
 CC enzymatic synthesis of peptides. They have enhanced stability towards
 CC oxidation under alkaline conditions, e.g. towards bleaching agents of the
 CC peroxy type. The invention also provides mutant cells in which the
 CC protease activity is diminished. Such cells can be used for the
 CC production of heterologous recombinant proteins
 XX
 SQ Sequence 636 AA;
 Query Match 88.5%; Score 1989.5; DB 2; Length 636;
 Best Local Similarity 87.3%; Pred. No. 1e-139;
 Matches 379; Conservative 30; Mismatches 24; Indels 1; Gaps 1;
 QY 1 NDVARGIVKADVAQSSVGLYGQGIIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
 DB 204 NDVARGIVKADVAQNNYGLYGQQLVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 263
 QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNLQTLFQAYS 120
 DB 264 NASDPNGHGHVAGSVLGN-ALNKGMAPOANLVFQSIMDSGGGLGSLPSNLQTLFQAWN 322
 QY 121 AGARIHTNSGAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 DB 323 AGARIHTNSGAPVNGAYTANSGQVDEYVRNNDMTILFAAGNEGPNSGTISAPGTAKNAI 382
 QY 181 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIKPDVMAPTGTFILSARSSLAPDSF 240
 DB 383 TVGATENYRPSFGSIADNPNIHQFSSRGATRDGRIKPDVTPAGTFILSARSSLAPDSF 442
 QY 241 WANHDSKYAMGTSMTATPIVAGNVAQLREHFVNRGKITPKPSLLKAALIAAGADIGLY 300
 DB 443 WANYSKYAYMGTSMTATPIVAGNVAQLREHFVNRGKITPKPSLLKAALIAAGADVGLGY 502
 QY 301 PNGNQGWGRVTLDKSLNVAYNVNESSLSSTQKATYSFTATAGKPLKISLVWSDAPASTTA 360
 DB 503 PNGDQGWGRVTLDKSLNVAYNVNEATATGQKATYSFQAQAGKPLKISLVWTDAPGSTTA 562
 QY 361 SVTLVNDLVLITAPNGTQVGNDFSPYNDWGRNENVENFINAPQSGTYYIEVQAYN 420
 DB 563 SYTLVNDLVLITAPNGQKYVGNDFSAFYDNNWGRNENVENFINAPQSGTYYIEVQAYN 622
 QY 421 VPVGPOTFSLAIYN 434
 DB 623 VPSGPORFSLAIHV 636
 RESULT 15
 ID AAM50083
 XX AAM50083 standard; protein; 433 AA.

AC AAM50083;
 XX 12-AUG-2002 (first entry)
 XX *Bacillus* sp Y-(FERM BP-1029) alkaline protease protein fragment.
 DE Alkaline protease; detergent; laundry; bleaching; dishwasher.
 XX *Bacillus* sp.
 XX EPI209233-A2.
 XX 29-MAY-2002.
 XX 22-NOV-2001; 2001EP-00127851.
 XX 22-NOV-2000; 2000JP-00355166.
 PR 12-APR-2001; 2001JP-00114048.
 XX (KAOS) KAO CORP.
 XX Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
 PI Okuda M, Saeki K;
 XX WPI; 2002-437518/47.
 DR New modified alkaline proteases useful in detergent compositions.
 XX Claim 5; Page 15-16; 25pp; English.
 XX This invention describes novel *Bacillus* sp. alkaline proteases useful in
 CC detergent compositions, especially in laundry, bleaching or automatic
 CC dishwasher detergents. The novel proteases have an increased detergency &
 CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
 CC sequence represents a fragment of the alkaline protease Ya from *Bacillus*
 CC sp strain Y-(FERM BP-1029) described in the method of the invention
 XX
 SQ Sequence 433 AA;
 Query Match 88.5%; Score 1987.5; DB 5; Length 433;
 Best Local Similarity 87.3%; Pred. No. 8.7e-140;
 Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;
 QY 1 NDVARGIVKADVAQSSVGLYGQGIIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
 DB 1 NDVARGIVKADVAQNNYGLYGQQLVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
 QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNLQTLFQAYS 120
 DB 61 NASDPNGHGHVAGSVLGN-ALNKGMAPOANLVFQSIMDSGGGLGSLPSNLQTLFQAWN 119
 QY 121 AGARIHTNSGAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 DB 120 AGARIHTNSGAPVNGAYTANSGQVDEYVRNNDMTILFAAGNEGPNSGTISAPGTAKNAI 179
 QY 181 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIKPDVMAPTGTFILSARSSLAPDSF 240
 DB 180 TVGATENYRPSFGSIADNPNIHQFSSRGATRDGRIKPDVTPAGTFILSARSSLAPDSF 239
 QY 241 WANHDSKYAMGTSMTATPIVAGNVAQLREHFVNRGKITPKPSLLKAALIAAGADIGLY 300
 DB 240 WANYSKYAYMGTSMTATPIVAGNVAQLREHFVNRGKITPKPSLLKAALIAAGADVGLGY 299
 QY 301 PNGNQGWGRVTLDKSLNVAYNVNESSLSSTQKATYSFTATAGKPLKISLVWSDAPASTTA 360
 DB 300 PNGDQGWGRVTLDKSLNVAYNVNEATATGQKATYSFQAQAGKPLKISLVWTDAPGSTTA 359
 QY 361 SVTLVNDLVLITAPNGTQVGNDFSPYNDWGRNENVENFINAPQSGTYYIEVQAYN 420
 DB 360 SYTLVNDLVLITAPNGQKYVGNDFSAFYDNNWGRNENVENFINAPQSGTYYIEVQAYN 419
 QY 421 VPVGPOTFSLAIYN 434
 DB 421 VPVGPOTFSLAIHV 434

Db 420 VPSGPORFSLAIVH 433

Search completed: March 10, 2004, 14:44:48
Job time : 61 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:43:43 ; Search time 22 Seconds
(without alignments)
1018.440 Million cell updates/sec

Title: US-09-985-689A-1
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSVGLY.....EVQAYNVVPGQTFSLAIVN 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2247	100.0	640	4	US-09-509-814A-6
2	2242	99.8	640	4	US-09-509-814A-8
3	2183	97.2	639	4	US-09-509-814A-4
4	2155	95.9	639	4	US-09-509-814A-1
5	2155	95.9	640	4	US-09-509-814A-2
6	2125.5	94.6	641	2	US-08-873-479-42
7	1986.5	88.4	433	4	US-09-104-623A-4
8	1986.5	88.4	433	4	US-09-019-532-4
9	1986.5	88.4	433	4	US-09-338-746-4
10	1986.5	88.4	635	2	US-08-873-479-43
11	1581.5	70.4	345	4	US-09-512-251A-10
12	1581.5	70.4	345	4	US-09-515-150A-10
13	1581.5	70.4	345	4	US-09-196-281-13
14	452.5	20.1	659	3	US-08-894-818B-1
15	452.5	20.1	659	4	US-09-445-472-12
16	414	18.4	412	4	US-09-445-472-1
17	414	18.4	522	3	US-08-894-818B-3
18	414	18.4	522	4	US-09-445-472-4
19	414	18.4	654	3	US-08-894-818B-35
20	414	18.4	654	4	US-09-445-472-16
21	401	17.8	659	3	US-08-894-818B-5
22	346	15.4	520	4	US-09-000-016-7
23	346	15.4	520	4	US-09-514-340-7
24	346	15.4	734	3	US-09-000-016-4
25	346	15.4	734	4	US-09-514-340-4
26	346	15.4	823	3	US-09-000-016-2
27	346	15.4	823	4	US-09-514-340-2

28	304.5	13.6	903	1	US-08-750-532-1	Sequence 1, Appli
29	304.5	13.6	1398	1	US-08-750-532-9	Sequence 9, Appli
30	304.5	13.6	1398	3	US-08-894-818B-8	Sequence 8, Appli
31	304.5	13.6	1398	4	US-09-445-472-6	Sequence 6, Appli
32	295.5	13.2	237	1	US-08-750-532-18	Sequence 18, Appli
33	283.5	12.6	418	2	US-08-873-479-44	Sequence 44, Appli
34	276	12.3	418	4	US-09-966-921A-2	Sequence 2, Appli
35	256.5	11.4	397	4	US-09-328-352-7533	Sequence 7533, Ap
36	254.5	11.3	275	1	US-08-431-387-1	Sequence 1, Appli
37	251	11.2	280	1	US-08-434-255-8	Sequence 8, Appli
38	251	11.2	280	1	US-08-459-967-8	Sequence 8, Appli
39	251	11.2	280	1	US-08-460-327-8	Sequence 8, Appli
40	251	11.2	280	1	US-08-459-871-8	Sequence 8, Appli
41	251	11.2	280	3	US-09-024-532-2	Sequence 2, Appli
42	251	11.2	280	4	US-09-104-623A-2	Sequence 2, Appli
43	251	11.2	280	4	US-09-019-532-2	Sequence 2, Appli
44	251	11.2	280	4	US-09-417-359A-2	Sequence 2, Appli
45	251	11.2	280	4	US-09-705-185-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-509-814A-6
; Sequence 6, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUM
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-OPCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 640
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-509-814A-6

Query Match 100.0%; Score 2247; DB 4; Length 640;
Best Local Similarity 100.0%; Pred. No. 3.le-173;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	NDVARGIVKADVAQSSVGLYGGQIVAVADTGLDGRDSSMHEAFRGKITALYALGRTN	60
Db	207	NDVARGIVKADVAQSSVGLYGGQIVAVADTGLDGRDSSMHEAFRGKITALYALGRTN	266
Qy	61	NANDTNGHGHVAGSVLGNSTKGMAPQANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS	120
Db	267	NANDTNGHGHVAGSVLGNSTKGMAPQANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS	326
Qy	121	AGARIHTNSWGAANVAYTDSRNVDYVRRKNDMTILFAAGNEGNGGTISAPGTAKNAI	180
Db	327	AGARIHTNSWGAANVAYTDSRNVDYVRRKNDMTILFAAGNEGNGGTISAPGTAKNAI	386
Qy	181	TVGATENTLRPSFGSYADNINHVAQSSRGPTKDGRIKPDVMAPGTIFLSARSLAPDSF	240
Db	387	TVGATENTLRPSFGSYADNINHVAQSSRGPTKDGRIKPDVMAPGTIFLSARSLAPDSF	446
Qy	241	WANHDSKYAYNGTSMATPIVAGNVAQLREHFVKRGITPKPSLLKAALIAGAADIGLGY	300

Db 447 WANHDSKYAMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 506
QY 301 PNGNQGWRTLDKSLNVAAYNSSLTSQKATYSFTATAGPKPLKISLVWSDAPASTTA 360
Db 507 PNGNQGWRTLDKSLNVAAYNSSLTSQKATYSFTATAGPKPLKISLVWSDAPASTTA 566
QY 361 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGNNVNFVFINAPQSGTYTIEVQAYN 420
Db 567 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGNNVNFVFINAPQSGTYTIEVQAYN 626
QY 421 VPVGPQTFSLAIVN 434
Db 627 VPVGPQTFSLAIVN 640

RESULT 2

US-09-509-814A-8

; Sequence 8, Application US/09509814A

; Patent No. 6376227

; GENERAL INFORMATION:

; APPLICANT: TAKAIWA, MIKIO

; APPLICANT: OKUDA, MITSUYOSHI

; APPLICANT: SAEKI, KATSUHIISA

; APPLICANT: KUBOTA, HIROMI

; APPLICANT: HITOMI, JUN

; APPLICANT: KAGEYAMA, YASUSHI

; APPLICANT: SHIKATA, SHITSUMI

; APPLICANT: NOMURA, MASAFUMI

; TITLE OF INVENTION: ALKALINE PROTEASE

; FILE REFERENCE: 0327-0832-0PCT

; CURRENT APPLICATION NUMBER: US/09/509,814A

; CURRENT FILING DATE: 2000-04-06

; PRIOR APPLICATION NUMBER: PCT/JP98/04528

; PRIOR FILING DATE: 1998-10-07

; PRIOR APPLICATION NUMBER: JP 9-274570

; PRIOR FILING DATE: 1997-06-08

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Patentin version 3.0

; SEQ ID NO 8

; LENGTH: 640

; TYPE: PRT

; ORGANISM: Bacillus sp.

US-09-509-814A-8

Query Match 99.8%; Score 2242; DB 4; Length 640;

Best Local Similarity 99.8%; Pred. No. 7.8e-173; Mismatches 1; Indels 0; Gaps 0;

Matches 433; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGTGRNDSMHEAPRGKITALYALGRTN 60
Db 207 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGTGRNDSMHEAPRGKITALYALGRTN 266
QY 61 NANDTNGHGHVAGSVLGNSTNGKMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 267 NANDTNGHGHVAGSVLGNSTNGKMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 326
QY 121 AGARIHTNSWGAAYVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db 327 AGARIHTNSWGAAYVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 386
QY 181 TVGATENLRPSFGSYADNINHVAFQSSRGPTKDGRIKPDVMAFGTILSARSLAPDSF 240
Db 387 TVGATENLRPSFGSYADNINHVAFQSSRGPTKDGRIKPDVMAFGTILSARSLAPDSF 446
QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
Db 447 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 506
QY 301 PNGNQGWRTLDKSLNVAAYNSSLTSQKATYSFTATAGPKPLKISLVWSDAPASTTA 360
Db 507 PNGNQGWRTLDKSLNVAAYNSSLTSQKATYSFTATAGPKPLKISLVWSDAPASTTA 566

QY 361 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGNNVNFVFINAPQSGTYTIEVQAYN 420
Db 567 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGNNVNFVFINAPQSGTYTIEVQAYN 626
QY 421 VPVGPQTFSLAIVN 434
Db 627 VPVGPQTFSLAIVN 640

RESULT 3

US-09-509-814A-4

; Sequence 4, Application US/09509814A

; Patent No. 6376227

; GENERAL INFORMATION:

; APPLICANT: TAKAIWA, MIKIO

; APPLICANT: OKUDA, MITSUYOSHI

; APPLICANT: SAEKI, KATSUHIISA

; APPLICANT: KUBOTA, HIROMI

; APPLICANT: HITOMI, JUN

; APPLICANT: KAGEYAMA, YASUSHI

; APPLICANT: SHIKATA, SHITSUMI

; APPLICANT: NOMURA, MASAFUMI

; TITLE OF INVENTION: ALKALINE PROTEASE

; FILE REFERENCE: 0327-0832-0PCT

; CURRENT APPLICATION NUMBER: US/09/509,814A

; CURRENT FILING DATE: 2000-04-06

; PRIOR APPLICATION NUMBER: PCT/JP98/04528

; PRIOR FILING DATE: 1998-10-07

; PRIOR APPLICATION NUMBER: JP 9-274570

; PRIOR FILING DATE: 1997-06-08

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Patentin version 3.0

; SEQ ID NO 4

; LENGTH: 639

; TYPE: PRT

; ORGANISM: Bacillus sp.

US-09-509-814A-4

Query Match 97.2%; Score 2183; DB 4; Length 639;

Best Local Similarity 96.3%; Pred. No. 4.5e-168; Mismatches 3; Indels 0; Gaps 0;

Matches 418; Conservative 13; Mismatches 3; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGTGRNDSMHEAPRGKITALYALGRTN 60
Db 206 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGTGRNDSMHEAPRGKITALYALGRTN 265
QY 61 NANDTNGHGHVAGSVLGNSTNGKMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 266 NANDTNGHGHVAGSVLGNSTNGKMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 325
QY 121 AGARIHTNSWGAAYVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db 326 AGARIHTNSWGAAYVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 385
QY 181 TVGATENLRPSFGSYADNINHVAFQSSRGPTKDGRIKPDVMAFGTILSARSLAPDSF 240
Db 386 TVGATENLRPSFGSYADNINHVAFQSSRGPTKDGRIKPDVMAFGTILSARSLAPDSF 445
QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
Db 446 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 505
QY 301 PNGNQGWRTLDKSLNVAAYNSSLTSQKATYSFTATAGPKPLKISLVWSDAPASTTA 360
Db 506 PNGNQGWRTLDKSLNVAAYNSSLTSQKATYSFTATAGPKPLKISLVWSDAPASTTA 565
QY 361 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGNNVNFVFINAPQSGTYTIEVQAYN 420
Db 566 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGNNVNFVFINAPQSGTYTIEVQAYN 625
QY 421 VPVGPQTFSLAIVN 434
Db 626 VPVGPQTFSLAIVN 639

```
RESULT 4
US-09-509-814A-1
Sequence 1, Application US/09509814A
Patent No. 6376227
GENERAL INFORMATION:
APPLICANT: TAKAIWA, MIKIO
APPLICANT: OKUDA, MITSUYOSHI
APPLICANT: SAEKI, KATSUHISA
APPLICANT: KUBOTA, HIROMI
APPLICANT: HITOMI, JUN
APPLICANT: KAGEYAMA, YASUSHI
APPLICANT: SHIKATA, SHITSUMI
APPLICANT: NOMURA, MASAFUMI
TITLE OF INVENTION: ALKALINE PROTEASE
FILE REFERENCE: 0327-0832-0PCT
CURRENT APPLICATION NUMBER: US/09/509, 814A
CURRENT FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: PCT/JP98/04528
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: JP 9-274570
PRIOR FILING DATE: 1997-06-08
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 639
TYPE: PRT
ORGANISM: Bacillus sp.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (23)..(23)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (29)..(29)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (32)..(32)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (46)..(46)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (47)..(47)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (53)..(53)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (70)..(70)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (74)..(74)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (89)..(89)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (102)..(102)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (105)..(105)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (128)..(128)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (130)..(130)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (131)..(131)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (132)..(132)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (133)..(133)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (146)..(146)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (148)..(148)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (160)..(160)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (165)..(165)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (172)..(172)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (183)..(183)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (187)..(187)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (188)..(188)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (189)..(189)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (194)..(194)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (286)..(286)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (306)..(306)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (324)..(324)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (369)..(369)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (431)..(431)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (501)..(501)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (531)..(531)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (541)..(541)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (584)..(584)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (591)..(591)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (592)..(592)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (594)..(594)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (595)..(595)
```

OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (396)..(596)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (611)..(611)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (632)..(632)
OTHER INFORMATION: Xaa is any amino acid
US-09-509-814A-1

Query Match 95.9%; Score 2155; DB 4; Length 639;
Best Local Similarity 96.3%; Pred. No. 8.3e-166;
Matches 418; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 1 NDVARGIVKADVAQSSYGLYGQIVAVADTGLDGRNDSSHEAFRCKITALVALGRTN 60
DB 206 NDVARGIVKADVAQSSYGLYGQIVAVADTGLDGRNDSSHEAFRCKITALVALGRTN 265
QY 61 NANDNGHGHVAGSVLNGSTNGKGMAPOLNVFOSIMDSGGGLGGLPSNLQTLFQAYS 120
DB 266 NANDNGHGHVAGSVLNGSTNGKGMAPOLNVFOSIMDSGGGLGGLPSNLQTLFQAYS 325
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGNGGTISAPGTAKNAI 180
DB 326 AGARIHTNSGAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGNGGTISAPGTAKNAI 385
QY 181 TVGATENLPSFGSVADNHNHVAQFSSRGPTKDGRIKPDVMAPTGFIILSARSSSLAPDSFF 240
DB 386 TVGATENLPSFGSVADNHNHVAQFSSRGPTKDGRIKPDVMAPTGFIILSARSSSLAPDSFF 445
QY 241 WANHSKYAIMGSTWATPIVAGNVAQREHFVVRKRGITPKPSLLKAALIGAADIIGLY 300
DB 446 WANHSKYAIMGSTWATPIVAGNVAQREHFVVRKRGITPKPSLLKAALIGAADIIGLY 505
QY 301 PNGNQGWGRTLDKSLNVAYNVNESSLSQATYSFTATAGKPLKISLWSDAPASTTA 360
DB 506 PNGNQGWGRTLDKSLNVAYNVNESSLSQATYSFTATAGKPLKISLWSDAPASTTA 565
QY 361 SVTLVNDLVLITAPNGTYGVNDFTSPYNDWGRNENNVFINAPSGTGTIVBQAYN 420
DB 566 SVTLVNDLVLITAPNGTYGVNDFTSPYNDWGRNENNVFINAPSGTGTIVBQAYN 625
QY 421 VPVGQTFSLAVN 434
DB 626 VPVGQTFSLAVN 639

RESULT 5
US-09-509-814A-2
Sequence 2, Application US/09509814A
Patent No. 6376227
GENERAL INFORMATION:
APPLICANT: TAKAIWA, MIKIO
APPLICANT: OKUDA, MITSUYOSHI
APPLICANT: SAEKI, KATSUHIKA
APPLICANT: KUBOTA, HIROMI
APPLICANT: HITOMI, JUN
APPLICANT: KAGEYAMA, YASUSHI
APPLICANT: SHIKATA, SHITSUMI
APPLICANT: NOMURA, MASAFUMI
TITLE OF INVENTION: ALKALINE PROTEASE
FILE REFERENCE: 0327-0832-OPCT
CURRENT APPLICATION NUMBER: US/09/509,814A
CURRENT FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: PCT/JP98/04528
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: JP 9-274570
PRIOR FILING DATE: 1997-06-08
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2

LENGTH: 640
TYPE: PRT
ORGANISM: Bacillus sp.
FEATURE:
NAME/KEY: misc feature
LOCATION: (3)..(3)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (24)..(24)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (30)..(30)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (33)..(33)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (47)..(47)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (48)..(48)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (54)..(54)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (71)..(71)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (75)..(75)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (90)..(90)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (103)..(103)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (106)..(106)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (129)..(129)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (131)..(131)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (132)..(132)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (133)..(133)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (134)..(134)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (147)..(147)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (149)..(149)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (161)..(161)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (166)..(166)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (173)..(173)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (184)..(184)
OTHER INFORMATION: Xaa is any amino acid

NAME/KEY: misc feature
LOCATION: (198)..(198)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (189)..(189)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (190)..(190)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (195)..(195)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (287)..(287)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (307)..(307)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (325)..(325)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (502)..(502)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (532)..(532)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (592)..(592)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (593)..(593)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (595)..(595)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (596)..(596)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (597)..(597)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (612)..(612)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (633)..(633)
OTHER INFORMATION: Xaa is any amino acid
US-09-509-814A-2

Query Match 95.9%; Score 2155; DB 4; Length 640;
Best Local Similarity 96.3%; Pred. No. 8.3e-166;
Matches 418; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 NDVARGIKADVASSYGLYGGQIVAVDTGLTGRNDSMHEAFRGKITIYALGRTN 60
DB 207 NDVARGIKADVASSYGLYGGQIVAVDTGLTGRNDSMHEAFRGKITIYALGRTN 266

QY 61 NANDTNHGTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
DB 267 NANDTNHGTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 326

QY 121 AGARIHTNSWGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
DB 327 AGARIHTNSWGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 386
QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVNMATPOTFILSARSSSLAPDSSF 240
DB 387 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVNMATPOTFILSARSSSLAPDSSF 446
QY 241 WANHDSKYAYMGTSMTPIVAGNVAQLREHFVNKRGITPKPSLLKALIAGAADIGLGY 300
DB 447 WANHDSKYAYMGTSMTPIVAGNVAQLREHFVNKRGITPKPSLLKALIAGAADIGLGY 506
QY 301 PNGNGWGRVTLDSKLNVAIVYNESSLSSTQKATYSFTATAGKPLKISLVWSDAPASTTA 360
DB 507 PNGNGWGRVTLDSKLNVAIVYNESSLSSTQKATYSFTATAGKPLKISLVWSDAPASTTA 566
QY 361 SVTLVNDLDAITAPNGTQYVGNDFTSYNDWMDGRNNVENVFNAPOSGTYTIEVQAYN 420
DB 567 SVTLVNDLDAITAPNGTQYVGNDFTSYNDWMDGRNNVENVFNAPOSGTYTIEVQAYN 626
QY 421 VVPGPQTFSLAIVN 434
DB 627 VVPGPQTFSLAIVN 640

RESULT 6
US-08-873-479-42
; Sequence 42, Application US/08873479
; Patent No. 5891701
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan
; APPLICANT: Lyne, Christianson
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
; TITLE OF INVENTION: Having Protease Activity
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5891701 No. 5891701disk of No. 5891701th America
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/873,479
; FILING DATE: 12-JUN-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Agrib, Cheryl H
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 5251.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 641 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-873-479-42

Query Match 94.6%; Score 2125.5; DB 2; Length 641;
Best Local Similarity 93.5%; Pred. No. 2e-163;
Matches 406; Conservative 19; Mismatches 8; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALVALGRTN 60
DB 209 NDVARGIVKADVAQNFGYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALVALGRTN 268
QY 61 NANDTNGHGHVAGSVLNGSTNKGWAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 269 NANDPENGHGHVAGSVLGN-ATNKGWAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 327
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
DB 328 AGARIHTNSGAPVNGAYTTDSRNVDYVVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 387
QY 181 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIKPDVMAQPTILSARSSSLAPDSSF 240
DB 388 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIKPDVMAQPTILSARSSSLAPDSSF 447
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLGY 300
DB 448 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLGY 507
QY 301 PNGNQGWGRVTLDKSLNVAYVNESSLSSTQKATYSTATAGPKPKISLVWSDAPASTTA 360
DB 508 PNGNQGWGRVTLDKSLNVAYVNESSLSSTQKATYSTATAGPKPKISLVWSDAPASTTA 567
QY 361 SVTLVNDLVLVITAPNGTQVGNDFTSYNDNDGRNNVNFVINAPOSCTYTIIEVOAYN 420
DB 568 SVTLVNDLVLVITAPNGTQVGNDFTSYNDNDGRNNVNFVINAPOSCTYTIIEVOAYN 627
QY 421 VPGVQPTFSLAIVN 434
DB 628 VPGVQPTFSLAIVH 641

RESULT 7

US-09-104-623A-4
; Sequence 4, Application US/09104623A
; Patent No. 6303752
; GENERAL INFORMATION:
; APPLICANT: Olsen, Arne Agerlin
; APPLICANT: Fatum, Tine Muxoll
; APPLICANT: Deussen, Heinz-Josef
; APPLICANT: Rosgen, Erwin Ludo
; TITLE OF INVENTION: A Modified Polypeptide
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6303752o No. 6303752disk of No. 6303752th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/104,623A
; FILING DATE: 25-JUN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 5256.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; STRANDEDNESS:

TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; STRAIN: Bacillus sp. Y
US-09-104-623A-4
Query Match 88.4%; Score 1986.5; DB 4; Length 433;
Best Local Similarity 87.3%; Pred. No. 1.9e-152; Indels 1; Gaps 1;
Matches 379; Conservative 29; Mismatches 25;
QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALVALGRTN 60
DB 1 NDVARGIVKADVAQNNGYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALVALGRTN 60
QY 61 NANDTNGHGHVAGSVLNGSTNKGWAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 61 NASDPENGHGHVAGSVLGN-ALNKGWAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAWN 119
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
DB 120 AGARIHTNSGAPVNGAYTTDSRNVDYVVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIKPDVMAQPTILSARSSSLAPDSSF 240
DB 180 TVGATENLRPSFGSIADNPNHIAQFSRSGATRGRIKPDVTAFTFILSARSLAPDSSF 239
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLGY 300
DB 240 WANVNSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLGY 299
QY 301 PNGNQGWGRVTLDKSLNVAYVNESSLSSTQKATYSTATAGPKPKISLVWSDAPASTTA 360
DB 300 PSGDQGWGRVTLDKSLNVAYVNEATLQCKATYSTFQAQAKPLKISLVMTDAPGSTTA 359
QY 361 SVTLVNDLVLVITAPNGTQVGNDFTSYNDNDGRNNVNFVINAPOSCTYTIIEVOAYN 420
DB 360 SVTLVNDLVLVITAPNGTQVGNDFTSYNDNDGRNNVNFVINAPOSCTYTIIEVOAYN 419
QY 421 VPGVQPTFSLAIVN 434
DB 420 VPGVQPTFSLAIVH 433

RESULT 8

US-09-019-532-4
; Sequence 4, Application US/09019532B
; Patent No. 6416756
; GENERAL INFORMATION:
; APPLICANT: Olsen, Arne Agerlin
; APPLICANT: Prent, Annette
; TITLE OF INVENTION: A Modified Enzyme for Skin Care
; FILE REFERENCE: 4922.204-US
; CURRENT APPLICATION NUMBER: US/09/019,532B
; CURRENT FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: 0038/97
; EARLIER FILING DATE: 1997-01-10
; EARLIER APPLICATION NUMBER: 0754/97
; EARLIER FILING DATE: 1997-06-25
; EARLIER APPLICATION NUMBER: 60/051,381
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: PCT/DK98/00015
; EARLIER FILING DATE: 1998-01-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-019-532-4

Query Match 88.4%; Score 1986.5; DB 4; Length 433;
Best Local Similarity 87.3%; Pred. No. 1.9e-152; Indels 1; Gaps 1;
Matches 379; Conservative 29; Mismatches 25;

```
QY 1 NDVARGIVKADVAGSSVGLYGGQGVAVADTGLDTGRNDSMHEAFRGKITALYALGRTN 60
DB 1 NDVARGIVKADVAGSSVGLYGGQGVAVADTGLDTGRNDSMHEAFRGKITALYALGRTN 60
QY 61 NANDTNGHGHVAGSVLNGSTNGKMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
DB 61 NASDPNGHGHVAGSVLNG-ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAWN 119
QY 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
DB 120 AGARIHTNSGAPVNGAYTANSRQVDEYVRNNDMTILFAAGNEGPNGGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADININHVAFSSRGPTKDGRIKPDVMAPOGTFFILSARSLAPDSSF 240
DB 180 TVGATENLRPSFGSIADPNPHIAQFSSRGATRDGRIKPDVTAPGTFFILSARSLAPDSSF 239
QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKRGTTPKPSLLKAALIAGAADIIGLY 300
DB 240 WANYSKYAYMGTSMATPIVAGNVAQLREHFVKRGTTPKPSLLKAALIAGATDVGLGY 299
QY 301 PNGNGWGRVTLDKSLNVAYNVNESSLSSTOKATYSFTATAGKPLKISLVKSDAPASTTA 360
DB 300 PSGDQGWGRVTLDKSLNVAYNVNEATATGOKATYSFOAQAGKPLKISLVWTDAPGSTTA 359
QY 361 SYTLVNDLVLITAPNGTQYVGNDFSTPYDNDMDGRNENVENFINAPQSGTYTIEVQAYN 420
DB 360 SYTLVNDLVLITAPNGQKYVGNDFSYPDNDMDGRNENVENFINAPQSGTYTIEVQAYN 419
QY 421 VPVGPQTFSLAIYN 434
DB 420 VPSGPQRFSLAIHV 433

RESULT 9
US-09-338-746-4
; Sequence 4, Application US/09338746
; Patent No. 6638526
; GENERAL INFORMATION:
; APPLICANT: Deussen, Heinz-Josef
; APPLICANT: Olsen, Arne A.
; APPLICANT: Fatum, Tine M.
; APPLICANT: Roggen, Erwin L.
; TITLE OF INVENTION: A Polypeptide-Polymer Conjugate
; FILE REFERENCE: 5619,200-US
; CURRENT APPLICATION NUMBER: US/09/338,746
; CURRENT FILING DATE: 1999-06-23
; EARLIER APPLICATION NUMBER: PA 1998 00809
; EARLIER FILING DATE: 1998-06-23
; EARLIER APPLICATION NUMBER: 60/091,461
; EARLIER FILING DATE: 1998-07-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-338-746-4

Query Match 88.4%; Score 1986.5; DB 4; Length 433;
Best Local Similarity 87.3%; Pred. No. 1.9e-152;
Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAGSSVGLYGGQGVAVADTGLDTGRNDSMHEAFRGKITALYALGRTN 60
DB 1 NDVARGIVKADVAGSSVGLYGGQGVAVADTGLDTGRNDSMHEAFRGKITALYALGRTN 60
QY 61 NANDTNGHGHVAGSVLNGSTNGKMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
DB 61 NASDPNGHGHVAGSVLNG-ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAWN 119
QY 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
DB 120 AGARIHTNSGAPVNGAYTANSRQVDEYVRNNDMTILFAAGNEGPNGGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADININHVAFSSRGPTKDGRIKPDVMAPOGTFFILSARSLAPDSSF 240
DB 180 TVGATENLRPSFGSIADPNPHIAQFSSRGATRDGRIKPDVTAPGTFFILSARSLAPDSSF 239
QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKRGTTPKPSLLKAALIAGAADIIGLY 300
DB 240 WANYSKYAYMGTSMATPIVAGNVAQLREHFVKRGTTPKPSLLKAALIAGATDVGLGY 299
QY 301 PNGNGWGRVTLDKSLNVAYNVNESSLSSTOKATYSFTATAGKPLKISLVKSDAPASTTA 360
DB 300 PSGDQGWGRVTLDKSLNVAYNVNEATATGOKATYSFOAQAGKPLKISLVWTDAPGSTTA 359
QY 361 SYTLVNDLVLITAPNGTQYVGNDFSTPYDNDMDGRNENVENFINAPQSGTYTIEVQAYN 420
DB 360 SYTLVNDLVLITAPNGQKYVGNDFSYPDNDMDGRNENVENFINAPQSGTYTIEVQAYN 419
QY 421 VPVGPQTFSLAIYN 434
DB 420 VPSGPQRFSLAIHV 433

RESULT 9
US-09-338-746-4
; Sequence 4, Application US/09338746
; Patent No. 6638526
; GENERAL INFORMATION:
; APPLICANT: Deussen, Heinz-Josef
; APPLICANT: Olsen, Arne A.
; APPLICANT: Fatum, Tine M.
; APPLICANT: Roggen, Erwin L.
; TITLE OF INVENTION: A Polypeptide-Polymer Conjugate
; FILE REFERENCE: 5619,200-US
; CURRENT APPLICATION NUMBER: US/09/338,746
; CURRENT FILING DATE: 1999-06-23
; EARLIER APPLICATION NUMBER: PA 1998 00809
; EARLIER FILING DATE: 1998-06-23
; EARLIER APPLICATION NUMBER: 60/091,461
; EARLIER FILING DATE: 1998-07-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-338-746-4

Query Match 88.4%; Score 1986.5; DB 4; Length 433;
Best Local Similarity 87.3%; Pred. No. 1.9e-152;
Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAGSSVGLYGGQGVAVADTGLDTGRNDSMHEAFRGKITALYALGRTN 60
DB 1 NDVARGIVKADVAGSSVGLYGGQGVAVADTGLDTGRNDSMHEAFRGKITALYALGRTN 60
QY 61 NANDTNGHGHVAGSVLNGSTNGKMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
DB 61 NASDPNGHGHVAGSVLNG-ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAWN 119
QY 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
DB 120 AGARIHTNSGAPVNGAYTANSRQVDEYVRNNDMTILFAAGNEGPNGGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADININHVAFSSRGPTKDGRIKPDVMAPOGTFFILSARSLAPDSSF 240
DB 180 TVGATENLRPSFGSIADPNPHIAQFSSRGATRDGRIKPDVTAPGTFFILSARSLAPDSSF 239
QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKRGTTPKPSLLKAALIAGAADIIGLY 300
DB 240 WANYSKYAYMGTSMATPIVAGNVAQLREHFVKRGTTPKPSLLKAALIAGATDVGLGY 299
QY 301 PNGNGWGRVTLDKSLNVAYNVNESSLSSTOKATYSFTATAGKPLKISLVKSDAPASTTA 360
DB 300 PSGDQGWGRVTLDKSLNVAYNVNEATATGOKATYSFOAQAGKPLKISLVWTDAPGSTTA 359
QY 361 SYTLVNDLVLITAPNGTQYVGNDFSTPYDNDMDGRNENVENFINAPQSGTYTIEVQAYN 420
DB 360 SYTLVNDLVLITAPNGQKYVGNDFSYPDNDMDGRNENVENFINAPQSGTYTIEVQAYN 419
QY 421 VPVGPQTFSLAIYN 434
DB 420 VPSGPQRFSLAIHV 433

RESULT 9
US-08-873-479-43
; Sequence 43, Application US/08873479
; Patent No. 5891701
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan
; APPLICANT: Lytne, Christianson
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
; TITLE OF INVENTION: Having Protease Activity
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5891701 No. 5891701disk of No. 5891701th America
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/873,479
; FILING DATE: 12-JUN-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Agis, Cheryl H.
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 5251,000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 635 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-873-479-43

Query Match 88.4%; Score 1986.5; DB 2; Length 635;
Best Local Similarity 87.3%; Pred. No. 3.3e-152;
Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAGSSVGLYGGQGVAVADTGLDTGRNDSMHEAFRGKITALYALGRTN 60
DB 203 NDVARGIVKADVAGSSVGLYGGQGVAVADTGLDTGRNDSMHEAFRGKITALYALGRTN 262
```



```
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Bacillus
US-09-196-281-13

Query Match      70.4%; Score 1581.5; DB 4; Length 345;
Best Local Similarity 95.3%; Pred. No. 7.1e-120;
Matches 303; Conservative 10; Mismatches 4; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSVGLYGQGIQVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Db 29 NDVARGIVKADVAQNNFGLYGQGIQVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 88
QY 61 NADNTNGHGHVAGSVLNGSTKNGWAPQANLVFQSIMDSGGGLGLPSNLQTLFSQAYS 120
Db 89 NADNPNGHGHVAGSVLNG-ATNKGWAPQANLVFQSIMDSGGGLGLPSNLQTLFSQAYS 147
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPGGTISAPGTAKNAI 180
Db 148 AGARIHTNSGAPVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPGGTISAPGTAKNAI 207
QY 181 TVGATENLRFSGSYADNINHVAQFSRGTGDKGRIPKPDVMAFGTILSARSSLAPDSF 240
Db 208 TVGATENLRFSGSYADNINHVAQFSRGTGDKGRIPKPDVMAFGTILSARSSLAPDSF 267
QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPILLSKAAIAGAADIGLY 300
Db 268 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPILLSKAAIAGAADVGLGF 327
QY 301 PNGNQGWGRVTLDKSLNV 318
Db 328 PNGNQGWGRVTLDKSLNV 345

RESULT 14
US-08-994-8183-1
; Sequence 1, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
```

```
; REFERENCE/DOCKET NUMBER: TAKAKURA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 659 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-894-818B-1

Query Match      20.1%; Score 452.5; DB 3; Length 659;
Best Local Similarity 30.1%; Pred. No. 2.5e-28;
Matches 138; Conservative 67; Mismatches 133; Indels 101; Gaps 18;

QY 8 VKADVAQSSYGLYGQGIQVAVADTGLDTRNDSSMHEAFRGKITALY-ALGRTNANDTN 66
Db 145 IGADTVNLSGLYDGGVVAIVDTGIDAN-----HPDLKGVIGWYDAVNGRSTPYDDQ 198
QY 67 GHGTHVAGSVLNGSTNK---GMAPOANLVFQSIM---DSGGGLGLPSNLQTLFSQAYS 121
Db 199 GHGTHVAGSVLNGSTNK---GMAPOANLVFQSIM---DSGGGLGLPSNLQTLFSQAYS 258
QY 122 GARI-----HTNSWGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPGGT 169
Db 259 GIRVNLISLGSQSSDGTSLSLQAVNNAWDA-----GIVVVAAGNSGPNITYT 306
QY 170 ISAPGTAKNAITVGATENLRFSGSYADNINHVAQFSRGTGDKGRIPKPDVMAFGTILS 229
Db 307 VGSFAAASKVITVGA-----VDSNDNIASFSSRGTADGRLEKPEVVAFGVDIIA 355
QY 230 ARSSLAPDSFPAWHDHSKYAYMGTSMATPIVAG-NVAQLREHFVKNRGITPK-PSLLK 286
Db 356 PRAS---GTSMTGTPINDYITKASGTSMTATPHVSGVGAIILOAH-----PSWTPDKVK 404
QY 287 AALIAGA-----ADIGLOYPNGNGWGRVTLDKSL---NVAYNSSSLSTSOKATY 335
Db 405 TALIETADIVAPKEIADIAGA-----GRVNYKAIKYDDYAKLFTGTVADKGSATH 457
QY 336 SFTATAGKPLKISLVMSDAPASTTASVTLVNDLVLITAPNGTOYVGNDFTSPYNDNWDG 395
Db 458 TFDVSGATFTATLYWD-----TGSSDIDLYLDPNNGNE-VDYSYTAYY----- 500
QY 396 RNNVENVFINAQSGTGTIEVOAYNVFVGPOTFSLAIVN 434
Db 501 --GFEKVGYNPTAGTWTGVVSYK---GAANYQVDVVS 534

RESULT 15
US-09-445-472-12
; Sequence 12, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMODO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 659
; TYPE: PRT
; ORGANISM: Thermococcus celer
US-09-445-472-12
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:46:23 ; Search time 39 Seconds
(without alignments)
2349.757 Million cell updates/sec

Title: US-09-985-689A-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSVGLY.....EVOAYNPVGPQTSLAIWN 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 21153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2247	100.0	434	10	US-09-985-689A-1
2	2247	100.0	434	15	US-10-385-662-2
3	2191	97.5	434	10	US-09-985-689A-2
4	2143	95.4	434	10	US-09-985-689A-6
5	2125.5	94.6	433	10	US-09-985-689A-7
6	1998.5	88.9	433	10	US-09-985-689A-5
7	1994.5	88.8	433	10	US-09-985-689A-3
8	1987.5	88.5	433	10	US-09-985-689A-4
9	1581.5	70.4	345	14	US-10-336-324-10
10	1581.5	70.4	345	14	US-10-403-105-13
11	452.5	20.1	659	13	US-10-090-624-12
12	414	18.4	412	13	US-10-090-624-1
13	414	18.4	522	13	US-10-090-624-4
14	414	18.4	654	13	US-10-090-624-16
15	366	16.3	1079	14	US-10-112-488-39

16	363.5	16.2	1208	14	US-10-156-761-13251	Sequence 13251, A
17	346.5	15.4	1139	14	US-10-156-761-10856	Sequence 10856, A
18	338	15.0	1237	14	US-10-314-657-4	Sequence 4, Appli
19	306.5	13.6	519	15	US-10-084-846A-114	Sequence 114, App
20	306.5	13.6	13725	15	US-10-084-846A-4	Sequence 4, Appli
21	304.5	13.6	1398	13	US-10-090-624-6	Sequence 6, Appli
22	283	12.6	580	10	US-09-927-827-55	Sequence 55, Appli
23	280.5	12.5	595	10	US-09-927-827-59	Sequence 59, Appli
24	276	12.3	418	9	US-09-966-921A-2	Sequence 2, Appli
25	270	12.0	1101	14	US-10-156-761-12934	Sequence 12934, A
26	251	11.2	280	14	US-10-209-812-2	Sequence 2, Appli
27	251	11.2	397	10	US-09-779-334A-5	Sequence 5, Appli
28	247	11.0	271	10	US-09-813-408-2	Sequence 2, Appli
29	246.5	11.0	379	10	US-09-813-408-6	Sequence 6, Appli
30	246	10.9	271	14	US-10-242-549-56	Sequence 56, Appli
31	244	10.9	271	14	US-10-242-549-46	Sequence 46, Appli
32	244	10.9	271	14	US-10-242-549-48	Sequence 48, Appli
33	244	10.9	271	14	US-10-242-549-50	Sequence 50, Appli
34	244	10.9	271	14	US-10-242-549-52	Sequence 52, Appli
35	244	10.9	271	14	US-10-242-549-54	Sequence 54, Appli
36	244	10.9	271	14	US-10-242-549-60	Sequence 60, Appli
37	242.5	10.8	627	10	US-09-927-827-60	Sequence 60, Appli
38	242	10.8	271	14	US-10-242-549-44	Sequence 44, Appli
39	242	10.8	271	14	US-10-242-549-58	Sequence 58, Appli
40	240	10.7	269	8	US-08-322-678-10	Sequence 10, Appli
41	240	10.7	269	9	US-09-837-235-16	Sequence 16, Appli
42	240	10.7	269	9	US-09-060-854B-6	Sequence 6, Appli
43	240	10.7	269	9	US-09-975-139-1	Sequence 1, Appli
44	240	10.7	269	9	US-09-976-414-8	Sequence 8, Appli
45	240	10.7	269	10	US-09-736-116-49	Sequence 49, Appli

ALIGNMENTS

RESULT 1

US-09-985-689A-1
; Sequence 1, Application US/09985689A
; Publication No. US2003002351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSURISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483US0
; CURRENT APPLICATION NUMBER: US/09/985,689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-1

Query Match 100.0%; Score 2247; DB 10; Length 434;
Best Local Similarity 100.0%; Pred. No. 6.2e-188;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	NDVARGIVKADVAQSSVGLYGGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN	60
DB	1	NDVARGIVKADVAQSSVGLYGGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN	60
QY	61	NANDTNGHGVAGSVLGNSTNKGMAPQANLVFQSIMDSGGGLGLPSNLQTLFSQAYS	120

```

Db      61  NNDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLPSNLQTLFSQAYS 120
QY      121  AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db      121  AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
QY      181  TVGATENLRSPSGSYADNINHVAQFSRSGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 240
Db      181  TVGATENLRSPSGSYADNINHVAQFSRSGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 240
QY      241  WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLGY 300
Db      241  WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLGY 300
QY      301  PNGNQGWGRVTLDKSLNVAIVNNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
Db      301  PNGNQGWGRVTLDKSLNVAIVNNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
QY      361  SVTLVNDLVLITAPNGTOYVGNDFTSFYNDNDGGRNNVENFINAPQSGTYYTIEVQAYN 420
Db      361  SVTLVNDLVLITAPNGTOYVGNDFTSFYNDNDGGRNNVENFINAPQSGTYYTIEVQAYN 420
QY      421  VPVGPOTFSLAIYN 434
Db      421  VPVGPOTFSLAIYN 434

```

RESULT 2

```

US-10-385-662-2
; Sequence 2, Application US/10385662
; Publication No. US20040002432A1
; GENERAL INFORMATION:
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: SAITO, KAZUHIRO
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: IZAWA, YOSHIFUMI
; APPLICANT: SAEKI, KATSUHISA
; APPLICANT: KOBAYASHI, TOHRU
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: Alkaline protease
; FILE REFERENCE: 234938US0
; CURRENT APPLICATION NUMBER: US/10/385,662
; PRIOR FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: JP 2002-081428
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: JP 2002-165987
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: JP 2002-304230
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: JP 2002-304231
; PRIOR FILING DATE: 2002-10-18
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Bacillus sp. KSM-KP43
US-10-385-662-2

```

```

Query Match      100.0%; Score 2247; DB 15; Length 434;
Best Local Similarity 100.0%; Pred. No. 6.2e-188;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1  NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRGNDSSMHEAFRGKITALYALGRTN 60
Db      1  NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRGNDSSMHEAFRGKITALYALGRTN 60
QY      61  NNDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLPSNLQTLFSQAYS 120
Db      61  NNDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLPSNLQTLFSQAYS 120

```

```

QY      121  AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db      121  AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
QY      181  TVGATENLRSPSGSYADNINHVAQFSRSGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 240
Db      181  TVGATENLRSPSGSYADNINHVAQFSRSGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 240
QY      241  WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLGY 300
Db      241  WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLGY 300
QY      301  PNGNQGWGRVTLDKSLNVAIVNNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
Db      301  PNGNQGWGRVTLDKSLNVAIVNNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
QY      361  SVTLVNDLVLITAPNGTOYVGNDFTSFYNDNDGGRNNVENFINAPQSGTYYTIEVQAYN 420
Db      361  SVTLVNDLVLITAPNGTOYVGNDFTSFYNDNDGGRNNVENFINAPQSGTYYTIEVQAYN 420
QY      421  VPVGPOTFSLAIYN 434
Db      421  VPVGPOTFSLAIYN 434

```

RESULT 3

```

US-09-985-689A-2
; Sequence 2, Application US/09985689A
; Publication No. US20030022351A1
; GENERAL INFORMATION: YUJI
; APPLICANT: HATADA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483US0
; CURRENT APPLICATION NUMBER: US/09/985,689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-14048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-2

```

```

Query Match      97.5%; Score 2191; DB 10; Length 434;
Best Local Similarity 96.5%; Pred. No. 4.9e-183;
Matches 419; Conservative 13; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1  NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRGNDSSMHEAFRGKITALYALGRTN 60
Db      1  NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRGNDSSMHEAFRGKITALYALGRTN 60
QY      61  NNDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLPSNLQTLFSQAYS 120
Db      61  NNDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLPSNLQTLFSQAYS 120
QY      121  AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db      121  AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
QY      181  TVGATENLRSPSGSYADNINHVAQFSRSGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 240
Db      181  TVGATENLRSPSGSYADNINHVAQFSRSGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 240

```

Qy 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIIGLY 300
Db 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIIGLY 300
Qy 301 PNGNQGWGRVTLDKSLNVAAYNNESSLSSTOKATYSTAGTKPLKISLVWSDAPASTTA 360
Db 301 PNGNQGWGRVTLDKSLNVAAYNNESSLSSTOKATYSTAGTKPLKISLVWSDAPASTTA 360
Qy 361 SVTLVNDLVLITAPNGTQYVGNDFSPYNDNDGRNNVENVFINAPQSGTYTIEVOAYN 420
Db 361 SVTLVNDLVLITAPNGTQYVGNDFSPYNDNDGRNNVENVFINAPQSGTYTIEVOAYN 420
Qy 421 VPVGPQTFSLAIVN 434
Db 421 VPVGPQTFSLAIVN 434

RESULT 4
US-09-985-689A-6
; Sequence 6, Application US/09985689A
; Publication No. US20030022351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483USO
; CURRENT APPLICATION NUMBER: US/09/985, 689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: JP P2001-114048
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-6

Query Match 95.4%; Score 2143; DB 10; Length 434;
Best Local Similarity 93.5%; Pred. No. 7, 7e-179;
Matches 406; Conservative 19; Mismatches 9; Indels 0; Gaps 0;

Qy 1 NDVARGIVKADVAQSSYGLYGGQGIIVADTGLDTGRNDSMHEAFRGKITAIYALGRTN 60
Db 1 NDVARGIVKADVAQSSYGLYGGQGIIVADTGLDTGRNDSMHEAFRGKITAIYALGRTN 60
Qy 61 NANDTNGHGHVAGSVLGNSTKGMAPAQANLVFQSIMDSGGGLGGLPSNLQTLFQOAYS 120
Db 61 NANDPNGHGHVAGSVLGNSTKGMAPAQANLVFQSIMDSGGGLGGLPSNLQTLFQOAYS 120
Qy 121 AGARIHTNSWGAANVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPGGTISAPGTAKNAI 180
Db 121 AGARIHTNSWGAANVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPGGTISAPGTAKNAI 180
Qy 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKPDYMAPGTFTLSARSSLAPDSSF 240
Db 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKPDYMAPGTFTLSARSSLAPDSSF 240
Qy 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIIGLY 300
Db 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIIGLY 300
Qy 301 PNGNQGWGRVTLDKSLNVAAYNNESSLSSTOKATYSTAGTKPLKISLVWSDAPASTTA 360
Db 301 PNGNQGWGRVTLDKSLNVAAYNNESSLSSTOKATYSTAGTKPLKISLVWSDAPASTTA 360

Db 301 PSNGQGWGRVTLDKSLNVAFVNETSSISTNOKATYSTAQSGKPLKISLVWSDAPASTTA 360
Qy 361 SVTLVNDLVLITAPNGTQYVGNDFSPYNDNDGRNNVENVFINAPQSGTYTIEVOAYN 420
Db 361 SVTLVNDLVLITAPNGTQYVGNDFSPYNDNDGRNNVENVFINAPQSGTYTIEVOAYN 420
Qy 421 VPVGPQTFSLAIVN 434
Db 421 VPVGPQTFSLAIVN 434

RESULT 5
US-09-985-689A-7
; Sequence 7, Application US/09985689A
; Publication No. US20030022351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483USO
; CURRENT APPLICATION NUMBER: US/09/985, 689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: JP P2001-114048
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-7

Query Match 94.6%; Score 2125.5; DB 10; Length 433;
Best Local Similarity 93.5%; Pred. No. 2, 6e-177;
Matches 406; Conservative 19; Mismatches 8; Indels 1; Gaps 1;

Qy 1 NDVARGIVKADVAQSSYGLYGGQGIIVADTGLDTGRNDSMHEAFRGKITAIYALGRTN 60
Db 1 NDVARGIVKADVAQSSYGLYGGQGIIVADTGLDTGRNDSMHEAFRGKITAIYALGRTN 60
Qy 61 NANDTNGHGHVAGSVLGNSTKGMAPAQANLVFQSIMDSGGGLGGLPSNLQTLFQOAYS 120
Db 61 NANDPNGHGHVAGSVLGN-ATNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFQOAYS 119
Qy 121 AGARIHTNSWGAANVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPGGTISAPGTAKNAI 180
Db 120 AGARIHTNSWGAANVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPGGTISAPGTAKNAI 179
Qy 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKPDYMAPGTFTLSARSSLAPDSSF 240
Db 180 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKPDYMAPGTFTLSARSSLAPDSSF 239
Qy 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIIGLY 300
Db 240 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIIGLY 299
Qy 301 PNGNQGWGRVTLDKSLNVAAYNNESSLSSTOKATYSTAGTKPLKISLVWSDAPASTTA 360
Db 300 PNGNQGWGRVTLDKSLNVAFVNETSPLSTOKATYSTAQSGKPLKISLVWSDAPGSTTA 359
Qy 361 SVTLVNDLVLITAPNGTQYVGNDFSPYNDNDGRNNVENVFINAPQSGTYTIEVOAYN 420
Db 360 SVTLVNDLVLITAPNGTQYVGNDFSPYNDNDGRNNVENVFINAPQSGTYTIEVOAYN 419
Qy 421 VPVGPQTFSLAIVN 434

RESULT 7
US-09-985-689A-3
; Sequence 3, Application US/09985689A

```
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483USO
; CURRENT APPLICATION NUMBER: US/09/985.689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-4

Query Match      88.5%; Score 1987.5; DB 10; Length 433;
Best Local Similarity 87.3%; Pred. No. 3e-165;
Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

Qy 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
Db 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
Qy 61 NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 61 NASDPNGHGHVAGSVLNG-ALNKGMAPQANLVFQSIIMDSGGGLGGLPSNLQTLFSQAWN 119
Qy 121 AGARIHTNSWGAAVNGAYTTDSRVNDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 120 AGARIHTNSWGAPVNGAYTANSRQVDEYVRNNDMTILFAAGNEGPNSTISAPGTAKNAI 179
Qy 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAFGTIFILSARSSLPDSSF 240
Db 180 TVGATENYRPSFGSIADPNPHIAQFSSRGATRDGRIKPDVTPGTFTILSARSSLPDSSF 239
Qy 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNGITPKPSLLKAALTAGAADIGLGY 300
Db 240 WANYNSKYAYMGTSMATPIVAGNVAQLREHFVKNGITPKPSLLKAALTAGADTVGLGY 299
Qy 301 PNGQGWGRVTLKSLNVAYNVNESSISTQKATYSTATAGKPLKISLVMSDAPASTTA 360
Db 300 PNGQGWGRVTLKSLNVAYNVEATATAGKATYSTQAGKPLKISLVMTDAPGSTTA 359
Qy 361 SVTLVNDLVLITAPNGTQVYNDFTSPYNDNWDGRNNVFNAPQSGTYTIEVOAYN 420
Db 360 SYTLVNDLVLITAPNGQKYVGNDFSPYDNNWDGRNNVFNAPQSGTYTIEVOAYN 419
Qy 421 VPGVGPQTFSLAIVN 434
Db 420 VPSGPGQRFSLAIVH 433

RESULT 9
US-10-336-324-10
; Sequence 10, Application US/10336324
; Publication No. US20030176304A1
; GENERAL INFORMATION:
; APPLICANT: Hansen, Peter
; APPLICANT: Bauditz, Peter
; APPLICANT: Mikkelsen, Frank
; APPLICANT: Andersen, Kim
; TITLE OF INVENTION: Protease Variants and Compositions
; FILE REFERENCE: 5349.204-US
; CURRENT APPLICATION NUMBER: US/10/336.324
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: US/09/512.251A
; PRIOR FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 10
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Bacillus
US-10-336-324-10

Query Match      70.4%; Score 1581.5; DB 14; Length 345;
Best Local Similarity 95.3%; Pred. No. 6.9e-130;
Matches 303; Conservative 10; Mismatches 4; Indels 1; Gaps 1;

Qy 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
Db 29 NDVARGIVKADVAQNNFGLYGQGVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 88
Qy 61 NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 89 NANDPNHGHVAGSVLGN-ATNKGMAPQANLVFQSIIMDSGGGLGGLPANLQTLFSQAYS 147
Qy 121 AGARIHTNSWGAAVNGAYTTDSRVNDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180

; SEQ ID NO 13
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Bacillus
US-10-403-105-13

Query Match      70.4%; Score 1581.5; DB 14; Length 345;
Best Local Similarity 95.3%; Pred. No. 6.9e-130;
Matches 303; Conservative 10; Mismatches 4; Indels 1; Gaps 1;

Qy 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
Db 29 NDVARGIVKADVAQNNFGLYGQGVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 88
Qy 61 NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 89 NANDPNHGHVAGSVLGN-ATNKGMAPQANLVFQSIIMDSGGGLGGLPANLQTLFSQAYS 147
Qy 121 AGARIHTNSWGAAVNGAYTTDSRVNDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
```

Db 148 AGARIHTNSWAPVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPGSCTISAPGTAKNAI 207
QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAFGTILSARSSSLAPDSF 240
Db 208 TVGATENLRPSFGSYADNINHVAFSSRGPTDGRIKPDVMAFGTILSARSSSLAPDSF 267
QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKRGITPSPSLKKAALIAAGADIGLY 300
Db 268 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKRGITPSPSLKKAALIAAGADVGLGF 327
QY 301 PNGNQGWGRVTLDKSLNV 318
Db 328 PNGNQGWGRVTLDKSLNV 345

RESULT 11
US-10-090-624-12
; Sequence 12, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 659
; TYPE: PRT
; ORGANISM: Thermococcus celer
US-10-090-624-12

Query Match 20.1%; Score 452.5; DB 13; Length 659;
Best Local Similarity 30.1%; Pred. No. 9.4e-31;
Matches 138; Conservative 67; Mismatches 153; Indels 101; Gaps 18;

QY 8 VKADVAQSSYGLYGQCIIVAVADTGLDTRNDSSMHEAFRGKITALLY-ALGRTNNDNTN 66
Db 145 IGADTWNLSLGYDGGVWVAIVDTGIDAN-----HPDLKGVICWYDAVNGRSTPYDDQ 198

QY 67 GHGTHVAGSVLNGSTNK---GMAPOANLVFQSIM--DSGGGLGLPSNLQTLFSAQYSA 121
Db 199 GHGTHVAGSVLNGSTNK---GMAPOANLVFQSIM--DSGGGLGLPSNLQTLFSAQYSA 121

QY 122 GARI-----HTNSWGAANVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPGGT 169
Db 259 GIRVINLSLGSQSSDGTDLSQLAVNNAWA-----GIVCVAGNSGPNYTT 306

QY 170 ISAPGTAKNAITVGAATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAFGTILS 229
Db 307 VGSPPAAASKVITVGA-----VDSNDNIAFSRSGPTADGRKLPKPEVAPGVADIIA 355

QY 230 ARSSLAPDSSFWANHDSKYAYMGTSMTATPIVAG-NVAQLREHFVKRGITPK--PSLLK 286
Db 356 PRAS---GTSMTGPIINDYTTAKSGTSMATPHVSGVGCALILQAH-----PSWTPDKVK 404

QY 287 AALIAGA-----ADIGLYPNGNQGWGRVTLDKSL---NVAYVNSSSLSTSQRATY 335
Db 405 TALLETADIVAPKEIADIAVGA-----GRNVYVKAIKYDDYAKLITFTGSVADKGSATH 457

QY 336 SFTATAGKPLKLSLWSDAPASTASVTLVNDLVLITAPNGTQYVGNDFTSYNDNMDGRN 395
Db 458 TFDVSGATFVATLYWD-----TGSSDIDLVLDPNGNE-VDYSYTAIY-----G 500

QY 396 RNNVENVFINAPSGTYTIEVQAYNVVPGPQTFSLAIVN 434
Db 501 --GFEKVGYNPTAGTWTVKVWSYK---GAANYQVDVWS 534

RESULT 12
US-10-090-624-1
; Sequence 1, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-10-090-624-1

Query Match 18.4%; Score 414; DB 13; Length 412;
Best Local Similarity 29.8%; Pred. No. 1.1e-27;
Matches 136; Conservative 59; Mismatches 147; Indels 114; Gaps 19;

QY 18 GLYGGQIIVAVADTGLDTRNDSSMHEAFRGKITALLYALGRTNAN-----DTNGHGH 71
Db 22 GYDGSGITIGITDID-----ASHPDLQGGK-----ICWVDFVNGRSYPYDDHGHGH 70

QY 72 VAGSVLNGSTN---KGMAPQANLVFQSIM--DSGGGLGLPSNLQTLFSAQYSAAGARI 125
Db 71 VASIAAGTGAASNGKYKGWAPGAKLAGIKVLGADGSGSISTIKGVEMAVDNKDKYGIKV 130

QY 126 HTNSWGA-----AVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPGGTISAP 173
Db 131 INLSLSSQSSDGTDLSQLAVNNAWA-----GLVWVAAGNSGPNKYTIGSP 178

QY 174 GTAKNAITVGAATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAFGTILSARSS 233
Db 179 AAASKVITVGA-----VDKVDVTSFSSRGPTADGRKLPKPEVAPGVADIIAARAS 227

QY 234 LAPDSSFWANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKRGITPK--PSLLKALIA 291
Db 228 ---GTSMGQPIINDYTTAAPTSGTSMATPHVAGIAALLLQ-----AHPSTWTPDKVKTALIE 277

QY 292 GA-----ADIGLYPNGNQGWGRVTLDKSLNVAAYVNSSSLSTSQA-----TYSFT 338
Db 278 TADIVKPEDEIADIAVGA-----GRNVAYKAIN--YDNYAKLVFTGYVANKGSQTHQFV 328

QY 339 ATAGKPLKLSLWSDAPASTASVTLVNDLVLITAPNGTQYVGNDFTSYNDNMDGRN 398
Db 329 ISGASFVATLYWDNAN-----SDLDLYLDPNGNQ-VDYSYTAIY-----G 369

QY 399 VENNVFINAPSGTYTIEVQAYNVVPGPQTFSLAIVN 434
Db 370 FEKVGYNPTDGTWTIKVWSYS---GSANYQVDVWS 402

RESULT 13
US-10-090-624-4
; Sequence 4, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: SHIMOJO, Tomoko
APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/10/090,624
CURRENT FILING DATE: 2002-03-06
PRIOR APPLICATION NUMBER: 09/445,472
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
PRIOR FILING DATE: 1997-06-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patent in version 3.0
SEQ ID NO 4
LENGTH: 522
TYPE: PRT
ORGANISM: Pyrococcus furiosus
FEATURE: misc feature
NAME/KEY: (428)..(428)
LOCATION: (428)..(428)
OTHER INFORMATION: Xaa at position 428 is Gly or Val.
US-10-090-624-4

Query Match 18.4%; Score 414; DB 13; Length 522;
Best Local Similarity 29.8%; Pred. No. 1.6e-27;
Matches 136; Conservative 59; Mismatches 147; Indels 114; Gaps 19;
QY 18 GLVGGQIVAVATGDTGRNDSMHEAFRGKITALYALGRTNAN-----DTNGHGTH 71
DB 22 GYDGGITIGIITGID-----ASHPDLOQKV-----IGWDFVNGRSYPYDDHGHGTH 70
QY 72 VAGSVLNGSTN---KGMAPQANLVFQSIM--DSGGGLGGLPSNLQTLFSQAYSAGARI 125
DB 71 VASTAAGTGAASNGKYKGMAPGAKLAGIKVLGADGSGSISTIIKGVWAVDNKDKYGIKV 130
QY 126 HTNSWGA-----AVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTISAP 173
DB 131 INLSGSSQSSDGTALSOAVNAAWDA-----GLVVVAAGNSGPNKYIIGSP 178
QY 174 GTAKNAITVGATENLRPSFGSYADNINHVAQFSRGPTKDGRIKPDYMAPGTFILSARSS 233
DB 179 AAASKVITVGA-----VDKYDVITSSRSRGPTADGRLKPEVVAPGNWIIAARAS 227
QY 234 LAPDSSFVANHDSKYAVMGSTMATPIVAGNVACLREHFVKNGRITPK--PSLLKAALIA 291
DB 228 --GTSMGQPINDIYTAAGTSMATPHVAGIAALLQ-----AHPSTPDKVKTALIE 277
QY 292 GA-----ADIGLYPNQNGRVTLDKSLNAVYVNESSLSLSQKA-----TYSFT 338
DB 278 TADIVKPEIADIAYGA-----GRVNAYKAIN--YDNYAKLVFTGYVANKGSOHQFV 328
QY 339 ATAGKPLKISLWSDAPASTTASVTLNDLVTAPNGTQYVGNDFTSYNDWNGRNN 398
DB 329 ISGASFVATLYWNNAN-----SDLDLYDPNGNQ--VDYSYTAY-----G 369
QY 399 VENVFIPAQSGTYTIEVQAYNVFVGPTFSLAIVN 434
DB 370 FEKUGYNNPTDGTWIKVWSYS---GSANYQVDVVS 402

RESULT 14
US-10-090-624-16
Sequence 16, Application US/10090624
Publication No. US2002013335A1
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: SHIMOJO, Tomoko
APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE

FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/10/090,624
CURRENT FILING DATE: 2002-03-06
PRIOR APPLICATION NUMBER: 09/445,472
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
PRIOR FILING DATE: 1997-06-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patent in version 3.0
SEQ ID NO 16
LENGTH: 654
TYPE: PRT
ORGANISM: Pyrococcus furiosus
US-10-090-624-16
Query Match 18.4%; Score 414; DB 13; Length 654;
Best Local Similarity 29.8%; Pred. No. 2.2e-27;
Matches 136; Conservative 59; Mismatches 147; Indels 114; Gaps 19;
QY 18 GLVGGQIVAVATGDTGRNDSMHEAFRGKITALYALGRTNAN-----DTNGHGTH 71
DB 154 GYDGGITIGIITGID-----ASHPDLOQKV-----IGWDFVNGRSYPYDDHGHGTH 202
QY 72 VAGSVLNGSTN---KGMAPQANLVFQSIM--DSGGGLGGLPSNLQTLFSQAYSAGARI 125
DB 203 VASTAAGTGAASNGKYKGMAPGAKLAGIKVLGADGSGSISTIIKGVWAVDNKDKYGIKV 262
QY 126 HTNSWGA-----AVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTISAP 173
DB 263 INLSGSSQSSDGTALSOAVNAAWDA-----GLVVVAAGNSGPNKYIIGSP 310
QY 174 GTAKNAITVGATENLRPSFGSYADNINHVAQFSRGPTKDGRIKPDYMAPGTFILSARSS 233
DB 311 AAASKVITVGA-----VDKYDVITSSRSRGPTADGRLKPEVVAPGNWIIAARAS 359
QY 234 LAPDSSFVANHDSKYAVMGSTMATPIVAGNVACLREHFVKNGRITPK--PSLLKAALIA 291
DB 360 --GTSMGQPINDIYTAAGTSMATPHVAGIAALLQ-----AHPSTPDKVKTALIE 409
QY 292 GA-----ADIGLYPNQNGRVTLDKSLNAVYVNESSLSLSQKA-----TYSFT 338
DB 410 TADIVKPEIADIAYGA-----GRVNAYKAIN--YDNYAKLVFTGYVANKGSOHQFV 460
QY 339 ATAGKPLKISLWSDAPASTTASVTLNDLVTAPNGTQYVGNDFTSYNDWNGRNN 398
DB 461 ISGASFVATLYWNNAN-----SDLDLYDPNGNQ--VDYSYTAY-----G 501
QY 399 VENVFIPAQSGTYTIEVQAYNVFVGPTFSLAIVN 434
DB 502 FEKUGYNNPTDGTWIKVWSYS---GSANYQVDVVS 534
RESULT 15
US-10-112-488-39
Sequence 39, Application US/10112488
Publication No. US20030082746A1
GENERAL INFORMATION:
APPLICANT: KIKUCHI, Yoshiaki
APPLICANT: DATE, Masayo
APPLICANT: UMEZAWA, Yukiko
APPLICANT: YOKOYAMA, Keiichi
APPLICANT: MATSUI, Hiroshi
TITLE OF INVENTION: PROCESS FOR PRODUCING TRANSGLUTAMINASE
FILE REFERENCE: 219286USOCNT
CURRENT APPLICATION NUMBER: US/10/112,488
CURRENT FILING DATE: 2002-04-01
PRIOR APPLICATION NUMBER: PCT/JP00/06780
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: JP2000-280098
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: JP11-280098
PRIOR FILING DATE: 1999-09-30
NUMBER OF SEQ ID NOS: 70

Search completed: March 10, 2004, 14:51:57
Job time : 40 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:42:47 ; Search time 21 Seconds
(without alignments)
1987.958 Million cell updates/sec

Title: US-09-985-689A-1
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSYGILY.....EVQAYNPVPGQTSLAIVN 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	523.5	23.3	1743	2 T18279	multidrug resist
2	497	22.1	1905	2 T18267	multidrug resist
3	347.5	15.5	444	2 B83891	intracellular alka
4	320.5	14.3	442	2 A69587	intracellular alka
5	310.5	13.8	806	2 A41341	microbial serine p
6	302.5	13.5	1398	2 T28159	pyrolysin (EC 3.4
7	283	12.6	580	2 S11890	serine proteinase
8	281	12.5	419	1 S25835	subtilisin (EC 3.4
9	280	12.5	799	2 G83753	subtilisin-type pr
10	279.5	12.4	1345	2 T29090	surface layer-asso
11	277	12.3	420	1 S23407	subtilisin (EC 3.4
12	267	11.9	715	2 JC4908	alkaline serine pr
13	263	11.7	1331	2 A72847	probable surface l
14	259.5	11.5	573	2 C84120	subtilisin-type pr
15	257.5	11.5	513	1 A35742	aqualysin (EC 3.4
16	253.5	11.3	894	2 F69730	cell wall-associat
17	251	11.2	627	2 D75393	serine proteinase,
18	246.5	11.0	402	1 JU0332	alkaline proteinase
19	246.5	11.0	534	1 J50173	alkaline proteinase
20	243	10.8	519	2 S71451	halolysin R4 (EC 3
21	242.5	10.8	401	2 I39974	serine proteinase
22	240	10.7	380	2 A49778	high-alkaline seri
23	240	10.7	488	2 A11930	proteinas import
24	235.5	10.5	382	1 SUBSN	subtilisin (EC 3.4
25	235	10.5	378	2 A33973	high-alkaline seri
26	235	10.5	382	2 I39780	subtilisin (EC 3.4
27	234	10.4	910	2 C69456	subtilisin sendai
28	234	10.4	1374	2 D72593	hypothetical prote
29	232.5	10.3	525	2 G84406	halolysin import

30	231	10.3	321	1 S27501	alkaline proteinase
31	230.5	10.3	379	1 SUBSCL	subtilisin (EC 3.4
32	229.5	10.2	601	2 JC4576	serine proteinase
33	227	10.1	1167	1 A35066	streptococcal Csa
34	226.5	10.1	1118	2 H97298	subtilisin like pr
35	225.5	10.0	381	2 JH0778	subtilisin (EC 3.4
36	225.5	10.0	613	2 S75976	hypothetical prote
37	225.5	10.0	1052	2 T17093	intraluminal subti
38	223.5	9.9	381	1 SUBSS	subtilisin (EC 3.4
39	223.5	9.9	381	1 SUBSI	subtilisin (EC 3.4
40	223.5	9.9	381	2 JQ1487	subtilisin (EC 3.4
41	222.5	9.9	275	2 JC1085	subtilisin (EC 3.4
42	222	9.9	384	2 JC4802	alkaline proteinase
43	221.5	9.9	272	2 A23624	subtilisin (EC 3.4
44	221.5	9.9	530	2 A42605	halolysin (EC 3.4
45	221	9.8	1036	2 JC5568	serine proteinase

ALIGNMENTS

RESULT 1

T18279
multidrug resistance transport protein - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
R:Shaulsky, G.; Loomis, W.F.
A:Accession: T18279
submitted to the EMBL Data Library, June 1996
A:Reference number: Z18855
A:Accession: T18279
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1743 <SHA>
A:Cross-references: EMBL:U60086; NID:g1399914; PID:g1399915; PIDN:AAB03331.1
C:Genetics:
A:Gene: tagC

Query Match 23.3%; Score 523.5; DB 2; Length 1743;
Best Local Similarity 27.9%; Pred. No. 2.3e-25;
Matches 165; Conservative 81; Mismatches 150; Indels 195; Gaps 24;

QY	19	LYGOGQIVAVADTGLDTR	-----NDS-----SMHEAFRGKITALYALGRNTNANDTNGH	68
DB	314	LRGKGQILSLADTGLDGHCFPSDSKYPILSNVNLNR	-----KVTVYITTTSDTDDSKVDGH	372
QY	69	GTHVAGSVLG	-----NGSTNKGMAFQANLVFQSLMDSGGGLGL--PSNLQTLFSQAY	119
DB	373	GTHICGSAAGTPEPDSVSNISFSGLATDAKIAF	-----FDLASGSSSLTPPSDLKQLYQPLY	429
QY	120	SAGARIHTNSGA	-----AVNGAYTDSRVDDYVRKN-DMTILFAAGNEGPNGTIS--A	172
DB	430	DAGARVHCDSWGVSVEGYTGSYSTASIDDFLTHPFIILRAAGN	-----NEQYLSLLT	486
QY	173	PGTAKNAITVGATENLR	-----PSFGSYADNI	199
DB	487	QSTAKNVTVGAHQTHENYLTDPGPNYINQSSVDINQELICDFSRVCNVTYTAQCLES	546	
QY	200	-----NHVAQSSRGPTKDGRIKPDVWAPGTFIL	228	
DB	547	NATTGLASCCPTLLRKSVIDAANTQPLLNNENNICSFSSKGPTHGDMKFPALVAPGEYIT	606	
QY	229	SARSSLA	-----PDSSFWANHDSKYAYMGGTSMATPIVAGNVAQLREH	272
DB	607	SARNGANTTDQCCDGLS-ENTNALLA-IGTSMATSFMAAAATILRQYLVDDGYVPTGSI	664	
QY	273	VKNRGITPKPSLLKAALIAAG	-----ADIGLGYPNGN	311
DB	665	VESNKLQPTGSLKALMINNAQLLNGTFLTSSSITYPSNQVFENFAGASLVQGWGAIR	724	
QY	312	LDSKNVAYVNES	-----SLSTSQKATYST--	338
DB	725	MSNHLHVNVNNNSNNNNKTSIDGITKFDGIGGLDLRLVKPNQWKEESLSTGNTSYCTFYK	784	

A;Residues: 1-442 <KUN>
A;Cross-references: GB:Z99113; GB:AL009126; NID:G2634090; PIDN:CAB13610.1; PID:el183385;
A;Experimental source: strain 168
C;Genetics:
A;Gene: aprX
C;Superfamily: subtilisin homology
F;146-398/Domain: subtilisin homology <SBT>

Query Match 14.3%; Score 320.5; DB 2; Length 442;
Best Local Similarity 29.7%; Pred. No. 3.5e-13;
Matches 107; Conservative 50; Mismatches 104; Indels 99; Gaps 16;

QY 11 DVAGSSGLYGGQGVAVADTGLDTRNDSSMEAFRGKITALYALGRTNAY-----D 64
DB EVVRNGQTGLGKGVAVADTGLDTRNDSSMEAFRGKITALYALGRTNAY-----D 183

QY 65 TNGHGTAVAGSVLNGSTN-----KGMAPQANLVFQSMDSGGGLGSLPNLQTLFQSAYS 120
DB EVVRNGQTGLGKGVAVADTGLDTRNDSSMEAFRGKITALYALGRTNAY-----D 183

QY 184 DNGHGTACGADVASSGSSGQYRGPAPEANLIGVKNKQGS-GTLADIIEGV----- 236
DB EVVRNGQTGLGKGVAVADTGLDTRNDSSMEAFRGKITALYALGRTNAY-----D 183

QY 121 AGARIHTNSGAAVN-----GAYTDSRVVDYVRKND-----MTIL 157
DB EVVRNGQTGLGKGVAVADTGLDTRNDSSMEAFRGKITALYALGRTNAY-----D 183

QY 158 FAAGNEPNGGTTISAPCTAKNAITVGTENLRPSFGSYADNINHVAFQSGSRGPTKDGRIK 217
DB EVVRNGQTGLGKGVAVADTGLDTRNDSSMEAFRGKITALYALGRTNAY-----D 183

QY 218 PDVMAPTFTILSRSSSLAPSSF-----WANHSKYAYMGTSMTATPIVAGNVQALREHF 272
DB EVVRNGQTGLGKGVAVADTGLDTRNDSSMEAFRGKITALYALGRTNAY-----D 183

QY 273 VKRGITPK--PSILKAALTAGADIGLYPNGNGQGRVTLDKSLNV-----AVNNESSL 327
DB EVVRNGQTGLGKGVAVADTGLDTRNDSSMEAFRGKITALYALGRTNAY-----D 183

QY 400 -QNPDLTPDEVKELK-----NGTDK-----KQEDENIYGAGVNAENSV 439
DB EVVRNGQTGLGKGVAVADTGLDTRNDSSMEAFRGKITALYALGRTNAY-----D 183

RESULT 5
A1341
Microbial serine proteinase (EC 3.4.21.-), minor (vpr), precursor - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 20-Jun-2000
C;Accession: A41341; B41341; S39700; D69730
R;Stroma, A.; Rufo Jr., G.A.; Theriault, K.A.; Dwyer, M.; Wilson, S.W.; Pero, J.
J. Bacteriol. 173, 6889-6895, 1991
A;Title: Cloning and characterization of the gene for an additional extracellular serine
A;Reference number: A41341; MUID:92041574; PMID:1938892
A;Accession: A41341
A;Molecule type: DNA
A;Residues: 1-806 <GLA>
A;Cross-references: GB:M76590; NID:G413819; PIDN:AAA2281.1; PID:G143820
A;Accession: B41341
A;Molecule type: protein
A;Residues: 161-195 <SL2>
R;Glaser, P.; Kunst, F.; Arnaut, M.; Coudart, M.P.; Gonzales, W.; Hullo, M.F.; Ionescu,
A.; Rapoport, G.; Danchin, A.
Mol. Microbiol. 10, 371-384, 1993
A;Title: Bacillus subtilis genome project: cloning and sequencing of the 97 kb region fr
A;Reference number: S39655; MUID:95020537; PMID:7934828
A;Accession: S39700
A;Molecule type: DNA
A;Residues: 1-806 <GLA>
A;Cross-references: EMBL:X73124; NID:G413923; PIDN:CAA51601.1; PID:G580871
R;Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillet, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabbet, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall
tech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Roettger, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
J.

A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
Akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpatra, P.; Togononi, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: D69730
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-806 <KUN>
A;Cross-references: GB:Z99123; GB:AL009126; NID:G2636240; PIDN:CAB15835.1; PID:G2636344
A;Experimental source: strain 168
C;Comment: The amino terminal sequence of the mature protein and a molecular weight of c
C;Genetics:
A;Gene: vpr
C;Superfamily: microbial serine proteinase vpr; subtilisin homology
C;Keywords: hydrolase; serine proteinase
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-160/Domain: propeptide #status predicted <PRO>
F;180-548/Domain: subtilisin homology #status atypical <SBT>

Query Match 13.8%; Score 310.5; DB 2; Length 806;
Best Local Similarity 23.5%; Pred. No. 3.4e-12;
Matches 129; Conservative 58; Mismatches 157; Indels 205; Gaps 18;

QY 18 GLYGQGVAVADTGLDTR-----NDSSMEAFRGKITALYALGRTN 60
DB GTGTGKIKVAIDTGVYHNHPLDKNFQYKGYGVNDVDPKPTPTG-----D 225

QY 61 NANDTNGHGTAVAGSVLNGSTNKGMAPQANLVFQSMDSGGGLGSLPNLQTLFQSAYS 120
DB PRGEATDGTGHTVAGTAAAG-TIKGVAPDALLAVRVLPQG--SGTTENVIAQVERAVQ 282

QY 121 AGARIHTNSGAAVNGAYTDSRVVDYVRKNDTILFAAGNEPNGGTTISAPCTAKNAI 180
DB DGADVMNLGLNSLNPDPWATSTAL-DWAMSEGVAVVTSNGNSGPNGTGSPGTSREAI 341

QY 181 TVGATE-----NLRFPSFGSY----- 195
DB SVGATQPLNEAVTVFGSYSSAKVNGYNKEDDVKALNNKVELVEAGIGEAKEKDFEGKDLT 401

QY 196 -----ADNI----- 199
DB 402 GKAVVYKRGSIATFADKADNNAKAGAGMNVYNNLSGSEIANVPGMVSPTIKLSLEDGKL 461

QY 200 -----NHVAFSSRGPTKD-GRIPDVMAPTFTILSRSSSLAP 236
DB 462 VSALKAGETKTFKLTVSKALGEQVADFSRGVPWMTWMIKPDISAPGVNIVSTIPTHDP 521

QY 237 DSSFWANHSKYAYMGTSMTATPIVAGNVQALREHFVKNRGITTPKPSL--LKAALIAGAA 294
DB 522 D-----HPYGYGSKQGTSMASPHIAGAVAIKQ-----AKPKWSVEQIKAAIMNTAV 568

QY 295 DI-----GLGYPNGNGWGRVTLDKSLNVAYVYNESSLSSTQKATYSFTATAGPKPLISLV 350
DB 569 TLKDSGGEVYHNHAGAGASAI-----NNALKAQSLVSPGYSY-----GTFLK----- 612

QY 351 WSDAPASTTASVTLVNDLVLITAPNGTQYQYGNDFTSYNDNDWDG--RNNVENVFINAPQ 408
DB 613 -ENGNETKNETFTIENQ-----SSIRKSYTLEYSFNGSGISTSGTSRVVIPAHQ 660

QY 409 SGTYYIEVQ 417
DB 661 TQKATAKVK 669

RESULT 6
T28159
Pyrolysin (EC 3.4.4.-) - Pyrococcus furiosus
C;Species: Pyrococcus furiosus
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 24-Oct-2000
C;Accession: T28159

R; Voorhorst, W.G.B.; Eggen, R.I.L.; Geerling, A.C.M.; Platteeuw, C.; Siezen, R.J.; Vos, J. Biol. Chem. 271, 20426-20431, 1996
 A:Title: Isolation and characterization of the hyperthermostable serine protease, pyroly
 A:Reference number: Z20481; MUID:96355370; PMID:8702780
 A:Accession: T28159
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1398 <VOO>
 A:Cross-references: EMBL:U55835; NID:g1556462; PID:g1556463; PIDN:AAB09761.1
 A:Experimental source: DSM3638
 C:Genetics:
 C:Keywords: hydrolase; serine proteinase

Query Match 13.5%; Score 302.5; DB 2; Length 1398;
 Best Local Similarity 26.6%; Pred. No. 2.3e-11;
 Matches 139; Conservative 57; Mismatches 178; Indels 149; Gaps 22;
 QY 21 GGGQIVAVADTGLDGRNDS-----SMHEAFRGKITALYALGRNNAN----- 63
 Db 301 GNGYDIYVDDTLDYDFDEVPLGQNVYDVAVSFYIGPLNYVLABIDPENGVEYAVFGW 360
 QY 64 DTNGHGTHVAGSLGNGSTN-----KGMAPQAN 91
 Db 361 DCHGHGTHVAGTVAGVDSNDAWDLMSYSGEVEFSRLYGWDYTNVTTDTVQGVAPGAQ 420
 QY 92 LVFQIMDSGGGLGSPNLQTLFQAYSAGARIHNSGAVNGAYT--TDSRN--VDD 147
 Db 421 INAIRVLRS--DGRGSMWDIEGM--TYAATHGADVISMISLGG--NAPYLDGTDPSVAVD 476
 QY 148 YVRKNDMTILFAAGNEFGNGGTISAPGTAKNAITVGTATENRPSFGSYAD----- 197
 Db 477 LTKYGVVFIAGNEGPGINIVSGFVATKALTVGAAA--VFINVGVVSQALGYEDYDG 535
 QY 198 -----NINVAOFSSRGPTKGRIPKDVWAPGTFILSARSSLAPSSFWANHDSKYA 249
 Db 536 FYFFPAYTNV-RIAPFSSRGPRIDGEIKPNVAVPGYGYSSLFWMWIGADF----- 585
 QY 250 YMGGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKALTAGAADI-----GLGYPNG 303
 Db 585 -MSGTSMATPHVSGVALLISG-PRPEGIYINPDIIKKVLSGATWLGSDPYTGQKYTEL 643
 QY 304 NQGWGRVTLKSLNVAYNSSLSSTQKATYFTATAGKPKLSLVWSDAPASTTA--- 360
 Db 644 DOGHGLVNVTKSWEI-----LKAINGTTLPIVDHWADKSYSDFAEYL 685
 QY 361 SYTLVNDLVLITAPN-----GTQYGN-----DFTSPYNDW-----DG-----RNNVNVF 403
 Db 686 GVDVIRGLVARNISIPDIVWEHVKYVDTEYRFEIYATEPWIKPFVSGSVILENNTEFVL 745
 QY 404 -----INAPQSGTY-----TIEVQAYNVVGPQTF 429
 Db 746 RVKYDVEGLEPLGVGRIIIDDPPTPVIEDEILNTIVIEKFT 788

RESULT 7
 S11890
 serine proteinase (EC 3.4.21.-) precursor, extracellular - Xanthomonas campestris pv. ca
 N:Alternate names: subtilisin-related proteinase
 C:Species: Xanthomonas campestris pv. campestris
 C:Date: 21-Nov-1993 #sequence_revision 07-Feb-1997 #text_change 03-Dec-1999
 C:Accession: S11890
 C:Jiu, Y.N.; Tang, J.L.; Clarke, B.R.; Dow, J.M.; Daniels, M.J.
 Mol. Gen. Genet. 220, 433-440, 1990
 A:Title: A multipurpose broad host range cloning vector and its use to characterise an e
 A:Reference number: S11890; MUID:90251253; PMID:2187155
 A:Accession: S11890
 A:Molecule type: DNA
 A:Residues: 1-580 <LIU>
 A:Cross-references: EMBL:X51635; NID:g48533; PIDN:CAA35962.1; PID:g48534
 A:Experimental source: Xanthomonas campestris pv. campestris
 A>Note: the sequence from Fig. 4 is inconsistent with that from Fig. 3 in having 205-Ala
 C:Superfamily: subtilisin; subtilisin homology

C:Keywords: extracellular protein; hydrolase; serine proteinase
 F:1-32/Domain: signal sequence #status predicted <SIG>
 F:168-423/Domain: subtilisin homology <SBT>

Query Match 12.6%; Score 283; DB 2; Length 580;
 Best Local Similarity 26.2%; Pred. No. 1.3e-10;
 Matches 127; Conservative 55; Mismatches 147; Indels 156; Gaps 24;

QY 21 GQQQIVAVADTGL-----DTGRNDSMHEAFRGKITALYALGRNNAN----- 64
 Db 168 GSGTVVAVIDTGITSHADLNANILAGYDFISDATTARDGNGRDSNAADGDMYAAECGA 227

QY 65 -----TNGHGTHVAGSLGNGSTNKGMAPOA-----NLVFGSIMD 99
 Db 228 GIPAASSWHGTHVAGTVAAVNTTGVAGTAGYAKVVPVVLGKGGSLDIADAIVWA 287

QY 100 SGGGLGGLPSNLQ--TLFSQAYSAGARIHNSGAAVNGAYTTDSRNVDYVRKNDMTIL 157
 Db 288 SGGTVSGIPANANPAEIVINMSLGGGSCSTTQON-AINGAVSRGT-----TVV 334

QY 158 FAAGNEGPGGTISAPGTAKNAITVGTATEN--LRPSFGSYADNINHVAFSSRGPTKGR 215
 Db 335 VAAGNDASNVSG--SLPANCANVIAAATTSAGAKASYSNFGTGI----- 377

QY 216 IKPDVWAPGTFILSARSS--LAPDSSFANHDSKYAYMGGTSMATPIVAGNVAQLREHFV 273
 Db 378 ---DVSAPOSSILSTLNSGTTTSGS-----ASYASYNGTSMASPHVAGVVALVQS--V 425

QY 274 KRGITPK--PSLLK--AALIAGAADILGYPNGOGWGRVTLKSLNVAYNSS----- 325
 Db 426 APTALTFAAVETLLKNTARALPGAC-----SGCGAGIVNADAATAA--INGSGGGG 477

QY 326 -----SLSTSQKATYSFTATAGKPLKISLVMSDAPASTTASVTL-----VND 367
 Db 478 GGGVTLTNGTPVTGLGAATCAELNYITVPAG-----SGILTVTTSGGSGD 523

QY 368 LDVLI---TAPNGTQVGNDFTPPYNDNWDGRNVENVFINAPQSGTYTIEVQAYNVFVG 424
 Db 524 ADLYVRAGSAPTDSAYT---CRPYRS-----GNAETCTITAP-SGTYYVRLKAYS----- 569

QY 425 PQTF 429
 Db 570 --TFS 572

RESULT 8

S25835
 subtilisin (EC 3.4.21.62) precursor - Bacillus sp. (strain TA41)

C:Species: Bacillus sp.
 C:Date: 22-Nov-1993 #sequence_revision 20-Feb-1995 #text_change 18-Jun-1999

C:Accession: S25835
 R:Davall, S.; Feiler, G.; Narinx, E.; Gerday, C.

Gene 119, 143-144, 1992

A:Title: Sequence of the subtilisin-encoding gene from an antarctic psychrotroph Bacill

A:Reference number: S25835; MUID:93012966; PMID:1398082

A:Accession: S25835

A:Molecule type: DNA

A:Residues: 1-419 <DAV>

A:Cross-references: EMBL:X63533; NID:g40198; PIDN:CAA45096.1; PID:g40199

C:Superfamily: subtilisin; subtilisin homology

C:Keywords: extracellular protein; hydrolase; serine proteinase

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-110/Domain: propeptide #status predicted <PRO>

F:111-419/Product: microbial serine proteinase #status predicted <MAT>

F:135-373/Domain: subtilisin homology <SBT>

F:144,184,359/Active site: Asp, His, Ser #status predicted

Query Match 12.5%; Score 281; DB 1; Length 419;
 Best Local Similarity 33.0%; Pred. No. 1.1e-10;
 Matches 87; Conservative 38; Mismatches 97; Indels 42; Gaps 13;

QY 21 GQQQIVAVADTGLDGRNDSMH-EAFRGKITALYALGRT---NNANDNGHGHVAGSV 76
 Db 21 GQQQIVAVADTGLDGRNDSMH-EAFRGKITALYALGRT---NNANDNGHGHVAGSV 76

Db 135 GAGINIAVLDTGVTNHPDLNNEVQCKD-----FTVGTNFTDSCDTRQGHGTHVAGSA 189
QY 77 LGNGSTNK---GMAPOANL-VFOSIMDSGGGLG-GLPSPNLQTLFQSAYSAGARIHTN-SW 130
Db 190 LANGGTGSGVYGVAPADLWAKVVGDDSGYADDAEIRHAGDQATALTNTKVNIMSL 249
QY 131 GAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNPGNGGTISAPGTAKNAITVGTATENLRP 190
Db 250 GSGGESSLIT---NAVDAIYDKGLVITAAAGNSGPKSGISYFGCALWNAVAVAALENTIQ 306
QY 191 SFGSYADNINHVAFQSSRGFTKDG-----RIKPDVMAPTFTLSARSSSLAPDSSFWANH 244
Db 307 N-GTY-----RVADFSSRGHRTAGDVYIQGDVFSAPGAAYST-----W-F 348
QY 245 DSKYVNGGTSMTPIVAGNVAQL 268
Db 349 DGGVATISGTSMAASHAAGLA 372
RESULT 9
G83753
subtilisin-type proteinase (EC 3.4.21.-) vpr precursor [similarity] - Bacillus halodurans
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: G83753
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A>Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: G83753
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-799 <SPT>
A:Cross-references: GB:AP001510; GB:BA000004; NID:gl0173440; PIDN:BA04550.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: vpr
C:Superfamily: microbial serine proteinase vpr; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:1-29/Domain: signal sequence #status predicted <SIG>
Query Match 12.5%; Score 280; DB 2; Length 799;
Best Local Similarity 24.7%; Pred. No. 3e+10;
Matches 117; Conservative 50; Mismatches 135; Indels 172; Gaps 16;
QY 18 GLXGQGIIVADVADTGLDTRGDSMHEAFRGKITALVALGRTNANDT-----NGH 68
Db 171 GYTGGITVALDTGVDTYTHPD--LVHAF-GDYKGWDFIDNDDPOETPPGDPRTGTH 227
QY 69 GTHVAGSVLNGSTNKGMAPQANLVFOSIMDSGGGLGSPNLQTLFQSAYSAGARIHTN 128
Db 228 GTHVAGTVAANGLT-KGVAPDANLLAYRVLPGG--RGSTAGVIAGTERAVQDQADIMNL 284
QY 129 SWGAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNPGNGGTISAPGTAKNAITVGTATENL 188
Db 285 SLGNTLNDPDPFATSIAL-DWMAEGVAVTISNGSGNNNTVGSPTGSRDAISVGAT--- 340
QY 189 RPSFGSY-----ADNINH----- 201
Db 341 RLPYKYNKASVFTSDGIDYPSADIMGPSPDELELDGETVEYAFAGLGRKPGDFGVDE 400
QY 202 -----VAQSSRGPT-KDGRIKPDVMAPTFTLSARSSSLAPD 237
Db 401 GKIALIVRGEIPFVEKAENAKAAGAVGAILIYNNVAGVQPTVPVGLAIPITMILNSHEDGLKMR 460
QY 202 -----VAQSSRGPT-KDGRIKPDVMAPTFTLSARSSSLAPD 237
Db 461 NELENGQNTVFTSIEFDKLVGETVADFSRGPVHWTWIKPDVSAFGVAIVSTPIPHQPD 520
QY 238 SSFWANHDSKYAVMGTSMTPIVAGNVAQLRHFVKNRGITPKPSLLKAALTAGADI- 296
Db 521 DPY-----GYGSRGTSMSAPHYVAGAAALLLEAH-PNWGV-----DHVKAALMNTAENLV 569

QY 297 ---GLGYPNGQGRVRLDKSLNAVYVNESSLSSTQKATY-SFTATAGPLK 346
Db 570 DENGNYRPHNTQGAG-----SIRVDAISETLVTPGSHSFGFTTKERGRQVE 617
RESULT 10
T29090
surface layer-associated STABLE proteinase - Staphylothermus marinus
N:Alternate names: hyperthermostable proteinase
C:Species: Staphylothermus marinus
C>Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T29090
R:May, J.; Lupas, A.; Kellermann, J.; Eckerskorn, C.; Baumeister, W.; Peters, J.
Curr. Biol. 6, 739-749, 1996
A>Title: A hyperthermostable protease of the subtilisin family bound to the surface layer
A:Reference number: Z20559; MUID:96385442; PMID:8793300
A:Accession: T29090
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1345 <MAY>
A:Cross-references: EMBL:U57968; NID:gl374755; PID:gl374756; PIDN:AA02323.1
A:Experimental source: strain F1
C:Function:
A:Description: probably serves an exodigestive function related to the organism's energy
A>Note: stoichiometric S-layer component
Query Match 12.4%; Score 279.5; DB 2; Length 1345;
Best Local Similarity 29.9%; Pred. No. 6.5e+10; Indels 81; Gaps 14;
Matches 99; Conservative 45; Mismatches 106; Indels 81; Gaps 14;
QY 46 FRGKITALVALGRTNANDTNGHGTVA-----GSVL-----GNGSTNK--GMAPQANLV 93
Db 445 YQGRYAL-----VSDFHGHGTSVATVIASGRVLDYLDGKLYRIMGVAPGAKI- 495
QY 94 PQSIMDSGGGLGSPNLQTLFQSAYSAG-----DYVRKNDMTILFAAGNPGNGGTISAPGTAK 177
Db 496 -----AGGDWLLGNLTLV--EAWLAGFNIVTEEDGYVYLSLDPGPHRADIIISNW 546
QY 131 GAAVNGAYTTDSRNV-----DVRKNDMTILFAAGNPGNGGTISAPGTAK 177
Db 547 GSIYINFWLQFPFGIDYRSSFMDDELAIARNYLIGHVTIVFAAGNPGNGGTISAPGTAK 606
QY 178 NAITVGATE--NLRPSFG---SYADNINHVAFQSSRGFTKDGRIKPDVMAPTFTLSARS 232
Db 607 LVITAGASTLMDYTRIYGYPEGYAD---EVIPFSSRGPTGQGYPKPDIVINIGAFEWASTR 663
QY 233 SLAPDSSFWANHDSKYAVMGTSMTPIVAGNVAQLRHFVKNRGITPKPSLLKAALIAI 292
Db 664 TI-DGRGVAQPD-----VFGTSEATPYTSGTLALIVQAYKEVNTTDPVTAKILLKSS 718
QY 293 AADIGLYPNGQGRVRLDKSLNAVYVNE 323
Db 719 AKDI--WYPAFSGSGRVDKALKAADTVFISE 747
RESULT 11
S23407
subtilisin (EC 3.4.21.62) 1 precursor - Bacillus sp. (strain TA39)
C:Species: Bacillus sp.
C>Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 18-Jun-1999
C:Accession: S23407
R:Narinx, E.; Davail, S.; Feller, G.; Gerday, C.
Biochim. Biophys. Acta 1131, 111-113, 1992
A>Title: Nucleotide and derived amino acid sequence of the subtilisin from the antarctic
A:Reference number: S23407; MUID:92256481; PMID:1581352
A:Accession: S23407
A:Molecule type: DNA
A:Residues: 1-420 <NAR>
A:Cross-references: EMBL:X62369; NID:G40200; PIDN:CAA44227.1; PID:G40201
C:Genetics:
A:Gene: subtilisin; subtilisin homology
C:Superfamily: extracellular protein; hydrolase; serine proteinase
C:Keywords: extracellular protein; hydrolase; serine proteinase

F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-111/Domain: propeptide #status predicted <PRO>
F;112-420/Product: microbial serine proteinase #status predicted <MAT>
F;136-374/Domain: subtilisin homology <SBT>
F;145,185,360/Active site: Asp, His, Ser #status predicted

Query Match 12.3%; Score 277; DB 1; Length 420;
Best Local Similarity 31.2%; Pred. No. 28-10;
Matches 98; Conservative 44; Mismatches 120; Indels 52; Gaps 16;

QY 21 GCGQIVAVADTGLDTRNDSSMHEAPRGKITAL--VALGRT---NKANDTNGHGHVAGS 75
DB 136 GCGINFAVLDTGVNTN-----HPDLRNNVEQCKDFTVGTITVNTNSCTDRQGHGHVAGS 189
QY 76 VLNGSTNK---GMAPQANL-VFQSIMDSGGGLG-GLPSNLTLSQAYSAGARHTN-S 129
DB 190 ALADGGTNGVGVAPADALWAYKVLGDDGSGYADIAAAIRHAGDQATLNTKVINMS 249
QY 130 WGAANVGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAITVGATENLR 189
DB 250 LGSSGESSLITNAV---YSYKGVLLIAAGNSGPGYQSGIGPGALYNVAVALEN-K 305
QY 190 PFGSVYADNINHAQPSRGPT-KDG-----RIKPDVWAPGTFILSARSLAPDSFWAN 243
DB 306 VENGTY-----RVADFSRGYSWTDGDAIQKGDVEISAPGAATYST-----W-- 348
QY 244 HDSKYAVMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKALIAAGADIGLYPNG 303
DB 349 FDGYATISGTSVASHAAGLAKIWAQYPSASNVDRVGLQYRAY---ENDILSGYAG 405
QY 304 -----NOGMRVTL 312
DB 406 YGDDFASGFGFATV 419

RESULT 12
JC4908
alkaline serine proteinase (EC 3.4.-.-) I precursor - Alteromonas sp.
N/Alternate names: subtilase
C/Species: Alteromonas sp.
C/Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 08-Oct-1999
C/Accession: JC4908
R/Tsujibo, H.; Miyamoto, K.; Tanaka, K.; Kaidzu, Y.; Imada, C.; Okami, Y.; Inamori, Y.
BioSci. Biotechnol. Biochem. 60, 1284-1288, 1996
A/Title: Cloning and sequence analysis of a protease-encoding gene from the marine bacterium *Alteromonas* sp.
A/Reference number: JC4908; MUID:97141200; PMID:8987544
A/Accession: JC4908
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-715 <TSU>
A/Cross-references: DDBJ:D38600; NID:91536787; PIDN:BAA18912.1; PID:dl019647; PID:g21602
A/Experimental source: strain O-7
C/Comment: This enzyme belongs to class I subtilisin-like family. It is a chelator-sensitized metalloprotease.
C/Genetics:
A/Gene: aprI
C/Superfamily: subtilisin homology
C/Keywords: hydrolase
F;1-40/Domain: signal sequence #status predicted <SIG>
F;41-150/Domain: amino-terminal propeptide #status predicted <ATP>
F;151-496/Product: alkaline serine protease I #status predicted <MAT>
F;182-452/Domain: subtilisin homology <SBT>
F;497-715/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F;239-294,335-372,478-481/disulfide bonds: #status predicted

Query Match 11.9%; Score 267; DB 2; Length 715;
Best Local Similarity 25.9%; Pred. No. 1.8e-09;
Matches 124; Conservative 51; Mismatches 168; Indels 136; Gaps 23;

QY 21 GCGQIVAVADTGLDTRNDSSMHEAPRGKITALYALGRTNAND-----TNG---- 67
DB 182 GCGVAVLDTGVRPHLDLNDANILPGYDMISNTFVANDGGARDNDARDGDAVTGECCT 241
QY 68 -----HGHVAG---SVLNGSTNKGMAPQANLVFQSIMDSGGGLGLP 108

DB 242 DSSGQPVPRADQSSWHGHVAGTAAVTNNGEGVAGVADAKVVPVRVL---GKCGGLT 298
QY 109 SNLQTLFSQAYSAGARIHTNSWGAIV-----NGAYTTDSRNVDDYVRKNDMTILFAAG 161
DB 299 SDIADGIIWASGGSDRVPANANPAVINMSLGGGACSAATTQAINQARNNGTVIVIAAG 358
QY 162 NEGPNGGTISAPGTAKNAITVGAT--ENLRPFGSVYADNINHAQPSRGPTDKGRIPKPD 219
DB 359 NNDNDSANYN-PGNCNGVNVVASVGRDGRAYYSNYGANI-----D 398
QY 220 VMAPGTFILSARSLAPDSFWANHDS-----KYAYMGTSMATPIVAGNVAQLR-- 269
DB 399 VAAPG---CAQSFADDPPEGILSTHNSGSGAPNSDSYHSQGTSMAPHVAGVAALIKQA 454
QY 270 -----BHFVN--RGITPKPSLLKALIAAGADI--GLG-----YPNGNGWGRVIL 312
DB 455 KPSATPDEVETILKNTTRSFAGSCNCGTGVVDAVAAVNEALGDVVTPPTGN-----TL 508
QY 313 DKSINVAVNVNESSLTSQKATYSFTATAGKPLKISLVMSDAPASTASVTL---VNDLD 369
DB 509 ED--GVAKTGLSAGASNQ--FTFDVPAGK-----TNVFTNSGGTGDAD 550
QY 370 LVITAPNGTYVGNDFTSFYNDNWGR---NNVNVFINAPQSGTYTIEVQAYNPVVG 424
DB 551 LYVKLL--GSQ-----PTSSSYDCRPYEGGNAEVCSEFDAPQAGTYHVMMINGYKAYS 599

RESULT 13
A72647
probable surface layer-associated STABLE proteinase APE0607 - Aeropyrum pernix (strain C/Species: Aeropyrum pernix
C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C/Accession: A72647
R/Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Taka-awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; DNA Res. 6, 83-101, 1999
A/Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aeropyrum* sp.
A/Reference number: A72450; MUID:99310339; PMID:10382966
A/Accession: A72647
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1331 <KAW>
A/Cross-references: DDBJ:AP000060; NID:95104188; PIDN:BAA79577.1; PID:dl043363; PID:g51
A/Experimental source: strain K1
C/Genetics:
A/Gene: APE0607

Query Match 11.7%; Score 263; DB 2; Length 1331;
Best Local Similarity 23.6%; Pred. No. 7.3e-09;
Matches 111; Conservative 65; Mismatches 158; Indels 136; Gaps 18;

QY 8 VKADVAQSSYGLYGGQGVAVADTGLDTRNDSSMHE-APRGKITALYA----- 55
DB 346 VYADLS-TAYVLF---LKALSDTGMISGPDPSLLDLSFADETPASYGVLEARDFTGD 400
QY 56 -----LGRTNAN-----DINGH 68
DB 401 GVNDPSAGALAGTYDWGLLTGESVNLGWRGLDFVAGLVPLGLDQGRWVSLYDTLAH 460
QY 69 GTHVAGSVLNGSTN-----KGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQA 118
DB 461 GTSVATVIASRGVFNGLGVITSLRGVAPGAKIA-----AGGSF-----LINVFAQL 509
QY 119 YSAG-----ARIHNSWG---AANVGAYT--TDSNVDDY-VRKNDMT 155
DB 510 FLSGFEPQDPLNWNVTGHEQVDVNNWSGNYSIALRGFLTGADDTATIDYIVSAGTV 569
QY 156 ILFAAGNEGNGGTISAPGTAKNAITVGATE--NLRPFGSVYADNINHAQPSRGPTKD 213
DB 570 IVHAWNGGFGYGTATTPGAGSLIISVGASTLFDYRFFGYLPSGCDVLSWSDRSPSQI 629
QY 214 GRIPDVWAPGTFILSARSLAPDSFWANHDSKYAVMGTSMATPIVAGNVAQLREHFV 273

```
Db 630 GVAKPDVNNIGSFANAG-----VPVLGLGNGSLAFDIFGTSSEATPWTSGSVALVISAYQ 685
Qy 274 KRGITPKPSLLKAALTAGAADIGLGYPNNGQMGWRVTLDKSLNV-----AY 320
Db 686 QAFGAKPSPLVKAALKSTARD*GA--DAFTQSGQGVVYRAVKAVLEGGVPALSTSVY 743
Qy 321 VNSSSLSTSQKATYSFTATAGKPKLSLWSDA--PASTASVTLVNDL 368
Db 744 ENVYSLLS-----GYSYFFLAAPNPVEDTQYPGVLKPGETAFTVTLVKTL 788

RESULT 14
C84120
subtilisin-type proteinase (EC 3.4.21.-) BHD763 precursor [similarity] - Bacillus halodurans
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: C84120
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: C84120
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-757 <STO>
A;Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BAB07482.1; GSPDB:GN00
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BR3763
C;Superfamily: microbial serine proteinase vpr; subtilisin homology
C;Keywords: hydrolase; serine proteinase
F;1-20/Domain: signal sequence #status predicted <SIG>

Query Match 11.5%; Score 259.5; DB 2; Length 757;
Best Local Similarity 22.6%; Pred. No. 5.7e-09;
Matches 119; Conservative 58; Mismatches 155; Indels 195; Gaps 20;

Qy 8 VKADVAQSSYGLYGGQIVAVADTGLTGRNDSMHEAFRG-----KITALYALG 57
Db 119 VRGMLDEGVHLTKGVKVAVIDTGIDYTHPD--LQSSYKGGVDYDYPDDPMETIASQ 176
Qy 58 RTNNANDTNGHTVAGSVLNGSTKMGAPQANLVFQSIIMDSGGGLGFLPSNLQTLFSQ 117
Db 177 -----PPTLHGTHVGIIAANGQV-KGVAPAEIYAYRALPGG--QGTEQVIAAEK 227
Qy 118 AYSAGARIHNSWGAAVNGAYTTDSRNVDYVAKNDMTILFAAGNEGPNNGTTISAPGTAK 177
Db 228 AVEDGVDVINLSLGTNVGPDWPTSLALDAAVEGVAVT-SNGNSGPNMTVGSFGTSK 286
Qy 178 NAITVGAT-----EN---LRPSFGS----- 194
Db 287 KAI SVGASAPPLNTPYLTAFCENEISLFPFSGGLPNAFKRDLPMIDVGVGTEKEWGD 346
Qy 195 -----YADINIH----- 201
Db 347 AEGKVVILKGMVFTKWHVAAXARGVLIYNNTPGPTGMEGGVNPVVSITREDG 406
Qy 202 -----VAOFSSRGP-TKDGRIKPDVMAFGTIFLSARSS 233
Db 407 EFLLEQLQKQKELTLRTYRKEEDFVALFSRGPVTHTWKPDVVAFGVSI-----DS 462
Qy 234 LAPDSSFWANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNGRITPK--PSLLKAALIA 291
Db 463 TIPNNG-----YLGNGTSMAPHVAGAAALIKO-----AHPEWTPEQVKALMN 507
Qy 292 GAADI-----GLGYPNNGQMGWRVTLDKSLNVAYVYNESSLSLTSQKATYSFTATAGKPKI 347
Db 508 TAKKLVDOEGVPHEIHEQGAGRIQVDRKAV-----AATSLVYPGALSFGK--- 551
Qy 348 SLVNSDAPASTASVTL-VNDLVLVITAPNGTVGVGNDFTSPYNDNW 393
Db 552 ---WSKODLREKRPVPTIENHDTV-----KRTYHISPPFDVPGVGV 591
```

```
RESULT 15
A35742
aqualysin (EC 3.4.21.-) I precursor - Thermus aquaticus
C;Species: Thermus aquaticus
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C;Accession: A35742; S00620; S00324
R;Kwon, S.T.; Terada, I.; Matsuzawa, H.; Ohta, T.
Eur. J. Biochem. 173, 491-497, 1988
A;Title: Unique precursor structure of an extracellular protease, aqualysin I, with NH-2
A;Reference number: A35742; MUID:90216674; PMID:2182621
A;Accession: A35742
A;Molecule type: DNA
A;Residues: 1-513 <TER>
A;Cross-references: GB:J90108; GB:D90108; GB:J05414; NID:g217171; PIDN:BAJ4135.1; PID:g
A;Note: the authors translated the codon CTG for residue 470 as Val, and GGT for residue
R;Kwon, S.T.; Terada, I.; Matsuzawa, H.; Ohta, T.
Eur. J. Biochem. 173, 491-497, 1988
A;Title: Nucleotide sequence of the gene for aqualysin I (a thermophilic alkaline serine
A;Reference number: S00620; MUID:88225062; PMID:3286255
A;Accession: S00620
A;Molecule type: DNA
A;Residues: 75-442 <KWO>
A;Cross-references: EMBL:X07734; NID:g48069; PIDN:CAA30559.1; PID:g602091
A;Note: part of this sequence, including the amino and carboxyl ends of the mature prote
R;Matsuzawa, H.; Tokugawa, K.; Hamachi, M.; Mizoguchi, M.; Taguchi, H.; Terada, I.; Kwon
Eur. J. Biochem. 171, 441-447, 1988
A;Title: Purification and characterization of aqualysin I (a thermophilic alkaline serin
A;Reference number: S00324; MUID:88151937; PMID:3162211
A;Accession: S00324
A;Molecule type: protein
A;Residues: 128-170 <MATS>
C;Superfamily: subtilisin; subtilisin homology
C;Keywords: extracellular protein; hydrolase; serine proteinase
F;1-14/Domain: signal sequence #status predicted <SIG>
F;15-127/Domain: propeptide #status predicted <PRO>
F;128-408/Product: aqualysin I #status experimental <MAT>
F;157-364/Domain: subtilisin homology <SBT>
F;255-257,281-283/Region: S1 specificity crevice #status predicted
F;409-513/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F;166,197,349/Active site: Asp, His, Ser #status predicted

Query Match 11.5%; Score 257.5; DB 1; Length 513;
Best Local Similarity 26.0%; Pred. No. 4.5e-09;
Matches 117; Conservative 44; Mismatches 144; Indels 145; Gaps 23;

Qy 16 SYGLYGGQIVAVADTGLTGRNDSMHEAFRGKITALY-ALGRTNANDTNGHGTAVG 74
Db 152 TVTATGRGVNYYVDTGIRT-----THREFGGRARVGYDALG--GNGQDCNGHGTAVG 203
Qy 75 SVLNGSTNGKMGAPQANLVFQSIIMD--SGGGLGFLPSNLQTLFQAYSAGARIHTN---- 128
Db 204 TI--GGVYTYGVAKAVNLAYRVFLDCNGSGTSGTGVAGVDWV-----TRNHRPFAVA 252
Qy 129 --SWGAAVNGAVTTDSRNVDYVRKN---DMTILFAAGNEGPNNGTTISAPGTAKNAITVG 183
Db 253 NMSLGGGVSTA-----LDNAVKNSTAGVYVAVAGNDNANACNYS-PARVAEALTVG 304
Qy 184 AT--ENLRPFSFGSYADNNHVAQSSRGPDKGRIKPDVMAFGTIFLSARSLAPDSFV 241
Db 305 ATTSSDARASFSNYGSCV-----DLFAPGASIFSA-----W 335
Qy 242 ANHDSKYAVMGTSMTATPIVAGNVAQLREHFVKNGRITP---KPSLLKAALIAAGADIGL 298
Db 336 YTSDFATQTLNGTSMTATPHVAG-VNAL--YLEQNPSTPASVASAILNGATGRLSGIGS 392
Qy 299 GYPNGNQMGWRVTLDKSLNVAYVYNESSLSLTSQKATYSFTATAGKPKLISLWSDAPAST 358
Db 393 GSPN-----RLLYSLSSGS-----GSTAPCTS 415
Qy 359 ----TASVTLVNDLVLVITAPNGTVG-----VGNFTSPYNDNWGRN---- 397
```


Db 416 CSYTCGSLSGPDYMF---QPNGTYYSFAGTHRAWLRGPACTDF-DLYLWRWDGSRWLT 471

Qy 398 -----NVENFFINAPQSGTYTIEVQAYN 420

Db 472 VGSSTGPTSEESLSYSGTAGYVLRRIYAYS 501

Search completed: March 10, 2004, 14:46:49
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:39:07 ; Search time 18 Seconds
(without alignments)
1255.469 Million cell updates/sec

Title: US-09-985-689A-1
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSGYL.....EQVQNVVPGVPTFSIAIVN 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	523.5	23.3	1743	1 TAGC DICTDI	Q23868 dictyosteli
2	497	22.1	1905	1 TAGB DICTDI	P54583 dictyosteli
3	310.5	13.8	806	1 SUBV_BACSU	P29141 bacillus su
4	304.5	13.6	1398	1 ELVA_BACSU	P27186 pyrococcus
5	283	12.6	580	1 EXPR_XANCP	P23314 xanthomonas
6	277	12.3	420	1 SUBT_BAC99	P28842 bacillus sp
7	257.5	11.5	513	1 AQLI_THEAQ	P08594 thermus aqu
8	253.5	11.3	894	1 WPRB_BACSU	P54423 bacillus su
9	246.5	11.0	402	1 ALP_CEPAC	P29118 cephalospor
10	246.5	11.0	534	1 PROA_VIBAL	P16588 vibrio albi
11	242.5	10.8	401	1 THES_BACSP	Q45870 bacillus sp
12	240	10.7	289	1 SUBS_BACLE	P29600 bacillus le
13	240	10.7	380	1 ELVA_BACAO	P27693 bacillus al
14	240	10.7	380	1 ELVA_BACCS	P41362 bacillus cl
15	239	10.6	269	1 PRMT_BACSP	Q99405 bacillus sp
16	235.5	10.5	382	1 SUBT_BACAM	P00782 bacillus am
17	235	10.5	378	1 ELVA_BACSP	P20724 bacillus sp
18	232	10.3	269	1 SUBB_BACLE	P29599 bacillus le
19	231	10.3	321	1 ISP_BACCS	P29140 bacillus cl
20	230.5	10.3	379	1 SUBT_BACLI	P00780 bacillus li
21	230	10.2	404	1 SMP1_MAGPO	Q9Y778 magnaporthe
22	230	10.2	1181	1 SCAL_STRPY	P58099 streptococ
23	227	10.1	1167	1 SCAL_STRPY	P15926 streptococ
24	225.5	10.0	381	1 SUBN_BACNA	P35835 bacillus su
25	225.5	10.0	1052	1 MS1P_CRIGR	Q92248 cricetulus
26	225.5	10.0	1052	1 MS1P_HUMAN	Q14703 homo sapien
27	225.5	10.0	1052	1 MS1P_MOUSE	Q9wtz2 mus musculu
28	225.5	10.0	1052	1 MS1P_RAT	Q9wtz3 rattus norv
29	223.5	9.9	381	1 SUBT_BACSA	P00783 bacillus su
30	223.5	9.9	381	1 SUBT_BACST	P29142 bacillus st
31	223.5	9.9	381	1 SUBT_BACSU	P04189 bacillus su
32	221.5	9.9	275	1 SUBT_BACPU	P07518 bacillus pu
33	221.5	9.9	530	1 HLY_HALI17	P29143 halophilic

ALIGNMENTS

RESULT 1

ID	TAGC DICTDI	STANDARD;	PRT;	1743 AA.
AC	Q23868;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Prestalk-specific protein tagC precursor (EC 3.4.21.-).			
GN	TAGC.			
OS	Dictyostelium discoideum (Slime mold).			
OC	Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.			
OX	NCBI_TaxID=44689;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=AX4;			
RX	MEDLINE=97140317; PubMed=986798;			
RA	Shaulsky G., Escalante R., Loomis W.F.;			
RT	"Developmental signal transduction pathways uncovered by genetic			
RT	suppressors".			
RL	Proc. Natl. Acad. Sci. U.S.A. 93:15260-15265(1996).			
CC	!- FUNCTION: Intercellular communication via tagC may mediate			
CC	integration of cellular differentiation with morphogenesis (By			
CC	similarity).			
CC	!- SIMILARITY: In the N-terminal section; belongs to peptidase family			
CC	S8.			
CC	!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING			
CC	PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.			
CC	!- SIMILARITY: STRONG, TO TAGS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U60086; AAB03331.1; --			
DR	PIR; T18279; T18279.			
DR	DictyBase; DDB0001795; tagC.			
DR	InterPro; IPR003593; AAA_ATPase.			
DR	InterPro; IPR001140; ABC_TM_transp.			
DR	InterPro; IPR003439; ABC_transporter.			
DR	InterPro; IPR002029; Peptidase_S8.			
DR	Pfam; PF00664; ABC membrane; 1.			
DR	Pfam; PF00005; ABC_tran; 1.			
DR	Pfam; PF00082; Peptidase_S8; 1.			
DR	PRINTS; PR00723; SUBTILISIN.			
DR	SMART; SM00382; AAA; 1.			
DR	PROSITE; PS50929; ABC_TM1F; 1.			
DR	PROSITE; PS0211; ABC_TRANSPORTER 1; 1.			
DR	PROSITE; PS50893; ABC_TRANSPORTER 2; 1.			
DR	PROSITE; PS00136; SUBTILASE ASP; FALSE_NEG.			
DR	PROSITE; PS00137; SUBTILASE_HIS; 1.			
DR	PROSITE; PS00138; SUBTILASE_SER; FALSE_NEG.			
KW	Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane;			

FT SIGNAL 1 31 POTENTIAL
FT CHAIN 32 1905 PRESTALK-SPECIFIC PROTEIN TAGB.
FT DOMAIN 378 700 PROTEASE.
FT DOMAIN 1518 1756 ABC TRANSPORTER.
FT TRANSMEM 1011 1031 POTENTIAL.
FT TRANSMEM 1076 1096 POTENTIAL.
FT TRANSMEM 1121 1141 POTENTIAL.
FT TRANSMEM 1210 1230 POTENTIAL.
FT TRANSMEM 1309 1329 POTENTIAL.
FT TRANSMEM 1332 1352 POTENTIAL.
FT ACT_SITE 387 387 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 432 432 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 695 695 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1553 1560 ATP (POTENTIAL).
FT DOMAIN 63 67 POLY-GLN.
FT DOMAIN 95 104 POLY-ASN.
FT DOMAIN 107 134 POLY-ASN.
FT DOMAIN 311 321 POLY-SER.
FT DOMAIN 833 837 POLY-SER.
FT DOMAIN 838 844 POLY-GLY.
FT DOMAIN 871 876 POLY-LEU.
FT DOMAIN 1012 1015 POLY-TLE.
FT DOMAIN 1385 1389 POLY-GLU.
FT DOMAIN 1398 1404 POLY-GLY.
FT DOMAIN 1445 1450 POLY-ASN.
FT DOMAIN 1765 1779 POLY-SER.
FT DOMAIN 1782 1785 POLY-SER.
FT DOMAIN 1807 1812 POLY-PRO.
FT DOMAIN 1813 1860 POLY-GLN.
FT DOMAIN 1872 1878 POLY-PRO.
FT CARBOHYD 594 594 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 672 672 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 747 747 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 823 823 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1172 1172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1522 1522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1658 1658 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1905 AA; 212518 MW; B8E223FA8B9AE13C CRC64;

Query Match 22.1%; Score 497; DB 1; Length 1905;
Best Local Similarity 28.0%; Pred. No. 1.6e-24;
Matches 162; Conservative 72; Mismatches 162; Indels 182; Gaps 22;

QY 19 LYQGOIVAVADGLDTGR---NDS-----SMHEAFRGKITALYALGRTNANDTNGH 68
DB 376 LRKGQGLLSIADTGLDGSCHFFSDSKYPIPFNQVNHENHRKVT---YIYHDNEDYVNGH 432
QY 69 GTHVAGSVLNG-----STKGMAPQANLVFQSDSGGLGLGPSNLQTLFQAYSA 121
DB 433 GTHVCGSAAGTPEDSSWAISPSGLATDAKIAFYD-LSSGSGSEPTPEPDSQMYXPLIDA 491
QY 122 GARIHNSWGA-----AVNGAYTDSNRVDYVRK-NDMTILFAAGNPGNGGTISAPGTA 176
DB 492 GARVHDSGMSVSLQGYGYSDDAGIDAFLEYEPEFSILRAAGN-NELFASLLAQATA 550
QY 177 KNAITVGATENLRPSFGS-----YADNI-----199
DB 551 KNAITVGAEGTAHVNVSDALEYDFSDNANFORPCLPKKVCNYTAKCCSEVSNVKG 610
QY 200 -----NHVAQFSRGTKDGRKPDVWAPGTILSARSS- 233
DB 611 QLCCPASIKONASDSFTTOPQFYNNENNGMSFSGKPTDGRKLPDIPVAPGEYITSARNSG 670
QY 234 -----LAPDSSFWANDSKYVAGGTSMTPIVAGNVAOLRHF-----VKNGRI 278
DB 671 ENSTDQGGDSL--PHANGILMSISGISMATPLATAATTLRLQVLVDGYPTGESVEENKL 728
QY 279 TPPELSLKALIAAGADIGLY-----PNGNGQGRVTRLDKSLNVA 319
DB 729 LPTGSLIKALMINNAQLNGTYFWSASSTNPSNAIPEQINGANLIGWGALRMN---NWL 785
QY 320 YNNESS-----SLSTSQKAT-----YSFT-----ATAGK 343

DB 786 YVKSNTPTSPRWIGIGLGNKQKATEWKEDSLSSGLNKSKYCTYKPKSSSSSGGGGT 845
QY 344 P-LKISLWSDAPASTTASVTLVNDL-----VITAPN--GTQYVGNFTSPYND 391
DB 846 PRIVATLVWTDPPSPYSYGAKEFNVLNVLNLLNSDDSIITIGNSGGSLQPAQKVAQP--- 902
QY 392 NWDGRNVNVFNAPQSGTYTTEVQAYNVVPGPQTES 429
DB 903 --DTLNVVEGIINPTKAMVKFTIAGTNVPIGPQKFS 938
RESULT 3
ID SUBV_BACSU STANDARD; PRT; 806 AA.
AC P29141,
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Minor extracellular protease vpr precursor (EC 3.4.21.-).
OS VPR OR IPA-45R OR BSU38090.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RP SEQUENCE FROM N.A., AND SEQUENCE OF 161-195.
RX MEDLINE=92041574; PubMed=1938892;
RA Sloma A., Rufo G.A. Jr., Theriault K.A., Dwyer M., Wilson S.W.,
Peto J.;
RT "Cloning and characterization of the gene for an additional
RT extracellular serine protease of Bacillus subtilis,"
RL J. Bacteriol. 173:6889-6895(1991).
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaud M., Condart M.P., Gonzales W.,
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees,"
RL Mol. Microbiol. 10:371-384(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.F.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghm S.Y., Glaser P., Goffeau A., Golightly E.J., Grand G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Heaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic C., Purnelle B., Rapoport G., Rev M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadie F.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccioni A., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzneger T.,


```

DR EMBL; AB012184; AAM40166.1; -.
DR PIR; S11890; S11890.
DR HSSP; P00782; 2SPT.
DR MEROPS; S08.00A; -.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR InterPro; IPR009020; Protease_inhib.
DR Pfam; PF04151; PPC; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
DR Hydrolase; Serine protease; Zymogen; Signal; Complete proteome.
KW SIGNAL 1 32 POTENTIAL.
FT PROPEP 33 7136 POTENTIAL.
FT CHAIN 7137 580 EXTRACELLULAR PROTEASE.
FT ACT_SITE 177 177 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 237 237 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 409 409 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 225 273 BY SIMILARITY.
FT DISULFID 315 352 POTENTIAL.
FT DISULFID 450 454 POTENTIAL.
SQ SEQUENCE 580 AA; 57228 MW; 8C9A2CAE4E7F47CB CRC64;

Query Match 12.6%; Score 283; DB 1; Length 580;
Best Local Similarity 26.2%; Pred. No. 3e-11;
Matches 127; Conservative 55; Mismatches 147; Indels 156; Gaps 24;

QY 21 GGGQIVAVADTGL----DTGRNDSMHEAFRGKITALYALGRTNNAND-----64
DB 168 GSGTVAVDTGITSHADINANILAGYDFISDATTARDGNGRDSNAEDGDWYAANECA 227
QY 65 -----TNGHGHVAGSLVNGSTNKGMAPQA-----NLVPSQIMD 99
DB 228 GTPASSWHGHVAGTVAAVNTTGTAGTAGYAKVPPVRLVGCGGSLSDIAATVWA 287
QY 100 SGGGLGSLNLQ--TLFSQAYSAGARIHTNSWGAANGAYTTDSRNVDDYVRKNDMTIL 157
DB 288 SGGTVSGIPANANPAEIVNMSLGGGSCSTTMON-AINGAVSRGT-----TVV 334
QY 158 FAAGNEGPGGTISAPGTAKNAITVGATEN--LRPSFGSYADINIHVAQPSRGTQGR 215
DB 335 VAAGNDASNVSG--SLPANCAVIAVAATTSAGAKASYNFGGI-----377
QY 215 IKPDVMAPTGTFILSRESS--LAPDSGFWANHDSKAYMGTSMATPIVAGNVAQLRHFV 273
DB 378 ----DVSGAPGSSILSTLNSGTTTPGS-----ASVASYNGTSTVSPHVGVALVQS--V 425
QY 274 KRGITPK--PSLLK--AALIGAADIGLYGNGQGRVTLDKSLNVAYNNESS-----325
DB 426 APTALTPAAVETLLKNTARALPGAC-----SGGCGAGIVNADAAVTA--INGSGGGG 477
QY 326 -----SLSTSQATYSFTATAGKPLKISLWSDAPASTTASVTL--VND 367
DB 478 GGGNTLTNGTPVTGLGAATGAELNYITVPAG-----SGTLTWTSGGSGD 523
QY 368 LDIVI---TAPNGTQVGNDFSTPYNDNDWGRNVNVEINAPQSGTYTIEVQAVNPG 424
DB 524 ADLYVRAGSAPTDASVT---CRPYRS-----GNAETCTITAP-SGTYVRLKAYS----569
QY 425 PQTES 429
DB 570 --TFS 572

RESULT 6
SUBT_BAC59 STANDARD; PRT; 420 AA.
AC P28842;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

```

```

DE Subtilisin precursor (SC 3.4.21.62).
GN SUB1.
OS Bacillus sp. (strain TA39).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=29336;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92256481; PubMed=1581352;
RA Narinx E., Davail S., Feller G., Gerday C.;
RT "Nucleotide and derived amino acid sequence of the subtilisin from
RL the Antarctic psychrotroph Bacillus TA39."
RL Biochim. Biophys. Acta 1131:111-113(1992).
CC -!- FUNCTION: Subtilisin is an extracellular alkaline serine protease,
CC it catalyzes the hydrolysis of proteins and peptide amides.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
CC for peptide bonds, and a preference for a large uncharged residue
CC in P1. Hydrolyzes peptide amides.
CC -!- COFACTOR: Binds 1 calcium ion per subunit (Potential).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: Still active at temperatures close to 0 degrees
CC Celsius, it has a marked heat lability.
CC -!- MISCELLANEOUS: Secretion of subtilisin is associated with onset of
CC sporulation, and many mutations which block sporulation at early
CC stages affect expression levels of subtilisin. However, subtilisin
CC is not necessary for normal sporulation.
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X62369; CAA44227.1; -.
DR PIR; S23407; S23407.
DR HSSP; Q99405; IMPT.
DR MEROPS; S08.00A; -.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
DR Hydrolase; Sporulation; Serine protease; Zymogen; Metal-binding;
KW Calcium-binding; Signal.
FT SIGNAL 1 26 POTENTIAL.
FT PROPEP 27 111 POTENTIAL.
FT CHAIN 112 420 SUBTILISIN.
FT ACT_SITE 145 145 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 182 182 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 360 360 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT METAL 115 115 CALCIUM (POTENTIAL).
FT METAL 154 154 CALCIUM (POTENTIAL).
SQ SEQUENCE 420 AA; 44086 MW; AE4F121BD32B26EC CRC64;

Query Match 12.3%; Score 277; DB 1; Length 420;
Best Local Similarity 31.2%; Pred. No. 4.8e-11;
Matches 98; Conservative 44; Mismatches 120; Indels 52; Gaps 16;

QY 21 GGGQIVAVADTGLDTGRNDSMHEAFRGKITAL--YALGRT---NNANDTNGHGHVAGS 75
DB 136 GGGINIAVLDGTGNTN-----HPDLRNVEQCKDFVTGTYTNNCTDRQGHGHVAGS 189
QY 76 VLNGSTNK---GMAPQANL-VFQSIMDSGGGLG-GLPSNLQTLFQAYSAGARIHTN-S 129
DB 190 ALADGGTGVGVGVPADLWYKVLGDDGSGVADDDIAAAIRHAGDQATALN*KKVINMS 249
QY 130 WGAANGAYTTDSRNVDDYVRKNDMTILPAAGNEGNGGTISAPGTAKNAITVGATENLR 189
DB 250 LGSSGSSSLITNAVN---YSYNGKVLIIAAAGNSGPGYQSGISGYPGALVNAVALEN-K 305

```

```

QY 190 PSFGSYADNINHAQESSRGPT-KDG-----RIKPDVMAPTFTILSARSSLPDSSFWAN 243
Db 306 VENGTT-----RVADSSGGYGTWDYAIQKGDVEISAFGAAYST-----W-- 348
QY 244 HDSKYAYMGTSMTATPIVAGNVAQREHFVKNRGITPKPSLLKAAALIAAGADIGYGPNG 303
Db 349 FDGGYATISGTSWASPHAAAGLAAKIAQWQPSASNDVVRGELQYRAY---ENDILSGYYAG 405
QY 304 -----NQGWGRVTL 312
Db 406 YGDFASFAGGFATV 419

RESULT 7
ID _AOL1_THEAQ STANDARD; PRT; 513 AA.
AC P08594;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aqualysin I precursor (EC 3.4.21.-).
GN PSTI.
OS Thermus aquaticus.
OC Bacteria; Deinococcus-Thermus; Thermus; Deinococci; Thermales; Thermaceae;
OC Thermus.
OX NCBI_TaxID=271;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 15-23.
RC STRAIN=VT1;
RX MEDLINE=90216674; PubMed=2182621;
RA Terada I., Kwon S.-T., Miyata Y., Matsuzawa H., Ohta T.;
RT "Unique precursor structure of an extracellular protease, aqualysin
RT I, with NH2- and COOH-terminal pro-sequences and its processing in
RT Escherichia coli."
RL J. Biol. Chem. 265:6576-6581(1990).
RN [2]
RP SEQUENCE OF 75-442 FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=VT1;
RX MEDLINE=88225062; PubMed=3286255;
RA Kwon S.-T., Terada I., Matsuzawa H., Ohta T.;
RT "Nucleotide sequence of the gene for aqualysin I (a thermophilic
RT alkaline serine protease) of Thermus aquaticus YT-1 and
RT characteristics of the deduced primary structure of the enzyme."
RL Eur. J. Biochem. 173:491-497(1988).
RN [3]
RP SEQUENCE OF 128-170.
RX MEDLINE=88151937; PubMed=3162211;
RA Matsuzawa H., Tokugawa K., Hamaoki M., Mizoguchi M., Taguchi H.,
RA Terada I., Kwon S.-T., Ohta T.;
RT "Purification and characterization of aqualysin I (a thermophilic
RT alkaline serine protease) produced by Thermus aquaticus YT-1."
RL Eur. J. Biochem. 171:441-447(1988).
CC -!- FUNCTION: Aqualysin I is a thermophilic alkaline serine protease.
CC The optimal temperature for its caseinolytic activity is 80
CC degrees Celsius.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DEVELOPMENTAL STAGE: Secreted from the early stationary phase
CC until the time the cells cease to grow.
CC -!- PTM: THE N- AND C-TERMINAL PRO-SEQUENCES ARE REMOVED THROUGH THE
CC PROTEOLYTIC ACTIVITY OF AQUALYSIN I ITSELF, IN THAT ORDER. THE
CC C-TERMINAL PRO-SEQUENCE IS REQUIRED FOR TRANSLOCATION OF THE
CC PROTEASES ACROSS THE OUTER MEMBRANE.
CC -!- PTM: Two disulfide bonds are present.
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

DR EMBL; D90108; BAA14135.1; -
DR EMBL; X07734; CAAG30559.1; -
DR PIR; A35742; A35742.
DR HSSP; P06873; 2PRK.
DR MEROPS; S08.051; -.
DR InterPro; IPR000209; Peptidase S8.
DR InterPro; IPR009020; Protease Inhib.
DR Pfam; PF00082; Peptidase S8; 1.
DR PRINTS; PRO0723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 14
FT PROPEP 15 127
FT CHAIN 128 408 AQUALYSIN I.
FT PROPEP 409 513
FT ACT SITE 166 166 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 197 197 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT SITE 349 349 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 513 AA; 53913 MW; DDFDFE6D4A50B785 CRC64;

Query Match 11.5%; Score 257.5; DB 1; Length 513;
Best Local Similarity 26.0%; Pred. No. 1.2e-09;
Matches 117; Conservative 44; Mismatches 144; Indels 145; Gaps 23;

QY 16 SYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALY-ALGRTNNANDTNGHGHVAG 74
Db 152 TYTATGGRVNVYVVDIGIRT-----THREFGGRVGVYDALG--GNGQDCNGHGHVAG 203
QY 75 SVLNGSTNGMAPQANLVFQSIQMD--SGGLGLPSNLQTLFQAYSAGARIHTN---- 128
Db 204 TI---GGVTYGVAKAVNLVAVRVLDCNGSGTSGVIAGVDWV-----TRNHRPAVA 252
QY 129 --SWGAAVNGAYTTDSRNVDDVYKRN---DMTILFAAGNEGNGGTISAPGTAKNAITVG 183
Db 253 NMSLGGGVST-----LDNAVKNIAAGVYVAVAGNDNANACNYS-PARVAEALTVG 304
QY 184 AT--ENLRPSFGSYADNINHAQESSRGPTKGRIPKPDVMAPTFTILSARSSLPDSSFW 241
Db 305 ATTSDDARASFSNYSVCV-----DLFAPGASIPSA-----W 335
QY 242 ANHDSKYAYMGTSMTATPIVAGNVAQREHFVKNRGITP---KPSLLKAAALIAAGADIGL 298
Db 336 YTSDTATQTLNGTSMATPHVAG-VAAL--YLEONPSATPASVASAILNGATTGRLSGIGS 392
QY 299 GYPNGNGGGRVTLDKSLNVAYVYVNESSLSISTSKATYSFTATAGKPLKISLVMSDPAST 358
Db 393 GSPN-----RLLYSLLSSGS-----GSTAPCTS 415
QY 359 ----TASVTLVNDLVLITAPNGTQY-----VGNDFTSPYNDVNDGRN--- 397
Db 416 CSYYTGLSLSGFDYNF---QPNGTYYSPAGTHRAWLRGPAGTDF-DLWLWRWDGSRWLT 471
QY 398 -----NVENVFINAPQSGTYTIEVOAYN 420
Db 472 VGSSTGFTSESLSYSGTAGYILWRIYAYS 501

RESULT 8
WPRA BACSU
ID _WPRA_BACSU STANDARD; PRT; 894 AA.
AC P54423; O06726;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE Cell wall-associated protease precursor (SC 3.4.21.-) [Contains: Cell
DE wall-associated polypeptides CWBP23 and CWBP52].
GN WPRA OR BSU10770.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
```


SEQUENCE FROM N.A., AND SEQUENCE OF 32-54 AND 414-428.
STRAIN=168;
MEDLINE=97159234; PubMed=9004506;
Margot P., Karamata D.;
"The wprA gene of *Bacillus subtilis* 168, expressed during exponential growth, encodes a cell-wall-associated protease.";
Microbiology 142:3437-3444(1996).
[2]
SEQUENCE FROM N.A.
STRAIN=168;
MEDLINE=98015415; PubMed=9353931;
Medina N., Vannier F., Roche B., Autret S., Levine A., Seror S.J.;
"Sequencing of regions downstream of addA (98 degrees) and citG (289 degrees) in *Bacillus subtilis*";
Microbiology 143:3305-3308(1997).
[3]
SEQUENCE FROM N.A.
STRAIN=168;
MEDLINE=98044033; PubMed=9384377;
Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G., Azavedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S., Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T., Entian K.D., Errington J., Fabret C., Ferrati E., Foulger D., Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N., Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G., Guisepi G., Guy B.J., Haga K., Hatched J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L., Joris B., Karamata D., Kasehara Y., Klaerr-Blanchard M., Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M., Kurita K., Lapius A., Lardinot S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Neback M., Neone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Paro V., Pohl T.M., Porcella D., Porwollik S., Prescott A.M., Presacan E., Pujic P., Purnelle D., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadale Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F., Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
"The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*";
Nature 390:249-256(1997).
-!- FUNCTION: NOT YET KNOWN; COULD BE INVOLVED IN PROTEOLYCAN DEGRADATION, BY CLEAVAGE OF ITS PEPTIDE BRIDGES.
-!- SUBCELLULAR LOCATION: Cell-wall bound.
-!- PTM: PROCESSED INTO CWBP23 AND CWBP52.
-!- SIMILARITY: Belongs to peptidase family S8.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch).

EMBL; U59891; AAC25926.1; -
EMBL; Y09476; CAA70641.1; -
EMBL; Z99109; CAB12917.1; -
PIR; F69730; F69730.
HSP; Q45670; LDB1.
MEROPS; S08.004; -
Subtilist; BG1846; wprA.
InterPro; IPR000209; Peptidase_S8.
Pfam; PF00082; Peptidase_S8; 1.

RT "Cloning and nucleotide sequences of the complementary and genomic
 RT DNAs for the alkaline protease from *Acremonium chrysogenum*.";
 RL Agric. Biol. Chem. 55:471-477(1991).
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; D0923; BAA0765.1; -
 DR PIR; J00332; J00332.
 DR HSSP; P06873; 2PRK.
 DR MEROPS; S08.UFA; -
 DR InterPro; IPR000209; Peptidase_S8.
 DR InterPro; IPR009020; Protease_inhib.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 DR KW Hydroxylase; Serine protease; Zymogen; Signal.
 FT SIGNAL 1 20 POTENTIAL.
 FT PROPEP 21 120 POTENTIAL.
 FT CHAIN 121 402 ALKALINE PROTEINASE.
 FT ACT_SITE 160 160 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 191 191 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 347 347 CHARGE RELAY SYSTEM (BY SIMILARITY).
 SQ SEQUENCE 402 AA; 42099 MW; 8D030CCD42D918E1 CRC64;
 Query Match 11.0%; Score 246.5; DB 1; Length 402;
 Best Local Similarity 30.3%; Pred. No. 4.4e-09;
 Matches 91; Conservative 32; Mismatches 102; Indels 75; Gaps 14;
 QY 21 GGGQIVAVADTGLDTRGNDSSMHEAFRGK-ITALYALGRITNNANDTNGHGHVAGSVLGN 79
 DB 151 GSGTYAYVVDGTI-----LESHNEFSGRAITGVNAVGGSN--ADINGHGHVAGTI--- 199
 QY 80 GSTNKGMAPQANLVFQSIMDSGGG-----LGGIPSNLQTLFQSNYSAGRI-----HT 127
 DB 200 GGRTYGVAKNTLAVKVFRRGSSSTSIILDGFNVAWVDIINRGONKAAISMSLGGVYS 259
 QY 128 NSWGAAYNGAYTTDSRVVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAITVG--AT 185
 DB 260 SAFNNAVNTAY--SRGVLSV-----RAGNDNQNAAYS-PASAANAIVTGSIAS 306
 QY 186 ENLRPFGSVADNINHVAQFSSRGPTKDGRIKPDVMAFGPTILSARSSILAPDSFWANHD 245
 DB 307 NWARSSFSNYGSVL-----DIFAPGTSILSA-----WIGGN 337
 QY 246 SKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADI---GLGYPN 302
 DB 338 SATNTISGTSMTATPHVTGVVLYLQ-----ALEGILTSGAARLALATTRVSNPFGSGSPN 393
 RESULT 10
 ID PROA VIBAL STANDARD; PRT; 534 AA.
 AC P16588;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Alkaline serine exoprotease A precursor (EC 3.4.21.-).
 GN PROA.
 OS Vibrio alginolyticus.
 OS Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 CC Vibrionaceae; Vibrrio.
 CX NCBI_TaxID=663;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=89326126; PubMed=2546861;
 RA Deane S.M., Robb F.T., Robb S.M., Woods D.R.;
 RT "Nucleotide sequence of the Vibrio alginolyticus calcium-dependent,
 RT detergent-resistant alkaline serine exoprotease A.";
 RL Gene 76:281-288(1989).
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M25499; AAA27550.1; -
 DR PIR; J00173; J00173.
 DR HSSP; Q99405; 1MPT.
 DR MEROPS; S08.050; -
 DR InterPro; IPR000209; Peptidase_S8.
 DR InterPro; IPR007280; PPC.
 DR InterPro; IPR009020; Protease_inhib.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR Pfam; PF04151; PPC; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 DR KW Hydroxylase; Serine protease; Zymogen; Signal.
 FT SIGNAL 1 21 POTENTIAL.
 FT PROPEP 22 141 POTENTIAL.
 FT CHAIN 142 534 ALKALINE SERINE EXOPROTEASE A.
 FT ACT_SITE 180 180 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 213 213 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 363 363 CHARGE RELAY SYSTEM (BY SIMILARITY).
 SQ SEQUENCE 534 AA; 55930 MW; 84E96D9C649D4226 CRC64;
 Query Match 11.0%; Score 246.5; DB 1; Length 534;
 Best Local Similarity 23.9%; Pred. No. 6.3e-09;
 Matches 114; Conservative 70; Mismatches 140; Indels 153; Gaps 24;
 QY 7 IVKADVAQSS--YGL-----YGGQIVAVADTGLDTRGNDSSMHEAF 46
 DB 137 IVSADANTATNATWGLDRIDQRNLPLDNNYSANFDTGTAYVIDTV-----NNAHVVF 190
 QY 47 RGTITALY-ALGRITNNANDTNGHGHVAGSVLGNSTNKGMAPQANLVFQSIM--DSGGG 103
 DB 191 GGRSVSGYDFVNDADADSCNGHGHVAGTI---GGSILGVAKXNVNLGVRLVSCSGGS 247
 QY 104 LGGIPSNLQTLFQSNYSAGRIHTNSWGAAYNGAYTTDSRVVDYVR---KNDMTILFAA 160
 DB 248 TSGVTIAGVDWAANA--SGPSVANMSLGGQSV-----LDSAVQSAVQSGVSYFMLAA 298
 QY 161 GNEGPNGGTISAPGTAKNAITVGAT--ENLRPFGSVADNINHVAQFSSRGPTKDGRIK 218
 DB 299 GNSNADACNYS-PARVATGVTVGSTTSTDARSSFNWGSV----- 338
 QY 219 DVMAFGPTILSARSSILAPDSFWANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRG 278
 DB 339 DVFAFGSQIKSA-----W--YDGGYKTSIGTSMATPHVAG-VAAL---YLQENSSV 383
 QY 279 TPKPSLLKAALIAAGADIGLVPGNQGWRVTLDKSLNAVAYVNESSSLTSQKATYSFT 338
 DB 384 S--PSQVEALIVRAST-----GKVT-----DTRGVNKLILSLT 416
 QY 339 -----ATAGKPL-----KISLVMSDAPASTASVTLLV---ND 367
 DB 417 DADCGQDCGGPDPTDPPEGLKTSGVFVSGSLGSSQVAYVYVDVEAGRLTVQWYGS 476
 QY 368 LDLVITAPNGTQVGNDETSPYNDWDQR-----NNVENVFINAQSGCTYITVEQAYN 420
 DB 477 ADLYLRFG-----AKPTLNAWDCRPFKYGNNETCTVSAQSGRIHYMIQSY 523

```

RESULT 11
THIS_BACSP STANDARD; PRT; 401 AA.
AC Q45670;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Thermophilic serine proteinase precursor (EC 3.4.21.-) (Ak.1
  protease).
OS Bacillus sp. (strain AK1).
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1409;
RN [1]_TaxID=1409;
RP SEQUENCE FROM N.A.
RX MEDLINE=95085262; PubMed=7993087;
RA Maciver B., McHale R.H., Saul D.J., Bergquist P.L.;
RT "Cloning and sequencing of a serine proteinase gene from a
  thermophilic Bacillus species and its expression in Escherichia
  coli.";
RL Appl. Environ. Microbiol. 60:3981-3988(1994).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.80 ANGSTROMS) OF 122-401.
RX MEDLINE=20057863; PubMed=10588904;
RA Smith C.A., Toogood H.S., Baker H.M., Daniel R.M., Baker E.N.;
RT "Calcium-mediated thermostability in the subtilisin superfamily: the
  crystal structure of Bacillus Ak.1 protease at 1.8-A resolution.";
RL J. Mol. Biol. 294:1027-1040(1999).
CC -|- COPACTOR: Binds 3 calcium ions and 1 sodium ion per subunit.
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- MISCELLANEOUS: Has a pH optimum of 8.5, a temperature optimum of
  75 degrees Celsius.
CC -|- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to licenses@isb-sib.ch).
CC -----
CC EMBL: L29506; AAA63688.1; -.
DR PIR: I39974; I39974.
DR PDB: 1DBI; 18-NOV-99.
DR MEROPS: S08.009; -.
DR InterPro: IPR000209; Peptidase_S8.
DR InterPro: IPR009020; Protease_inhib.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_SER; 1.
DR KW Hydrolyase; Serine protease; Zymogen; Metal-binding; Calcium-binding;
  KW Signal; 3D-structure.
FT SIGNAL 1 24 POTENTIAL.
FT PROPEP 25 121
FT CHAIN 122 401 THERMOPHILIC SERINE PROTEINASE.
FT ACT_SITE 160 160 CHARGE RELAY SYSTEM.
FT ACT_SITE 193 193 CHARGE RELAY SYSTEM.
FT ACT_SITE 347 347 CHARGE RELAY SYSTEM.
FT METAL 126 126 CALCIUM 1.
FT METAL 168 168 CALCIUM 3 (VIA CARBONYL OXYGEN).
FT METAL 169 169 CALCIUM 1.
FT METAL 171 171 CALCIUM 2.
FT METAL 179 179 CALCIUM 2.
FT METAL 184 184 CALCIUM 2.
FT METAL 186 186 CALCIUM 2 (VIA CARBONYL OXYGEN).
FT METAL 204 204 CALCIUM 1.
FT METAL 204 204 CALCIUM 3.
FT METAL 207 207 CALCIUM 1.
FT METAL 207 207 CALCIUM 1 (VIA CARBONYL OXYGEN).
FT METAL 209 209 CALCIUM 1 (VIA CARBONYL OXYGEN).
FT METAL 211 211 CALCIUM 1 (VIA CARBONYL OXYGEN).

```

```

297 297 297 METAL 297 SODIUM (VIA CARBONYL OXYGEN).
300 300 300 FT METAL 300 SODIUM (VIA CARBONYL OXYGEN).
323 323 323 FT METAL 323 SODIUM.
258 258 258 FT DISULFID
127 127 127 FT TURN
129 129 127 FT HELIX
132 132 132 FT TURN
135 135 135 FT TURN
137 137 137 FT HELIX
140 140 140 FT TURN
141 141 141 FT HELIX
144 144 144 FT TURN
145 145 145 FT TURN
152 152 153 FT TURN
155 155 160 FT STRAND
165 165 166 FT TURN
168 168 173 FT TURN
174 174 179 FT STRAND
180 180 183 FT TURN
184 184 184 FT STRAND
193 193 202 FT HELIX
219 219 224 FT STRAND
228 228 229 FT TURN
234 234 246 FT HELIX
247 247 248 FT TURN
251 251 254 FT STRAND
263 263 274 FT HELIX
275 275 276 FT TURN
278 278 282 FT STRAND
285 285 285 FT STRAND
298 298 299 FT TURN
301 301 306 FT STRAND
308 308 309 FT TURN
312 312 312 FT STRAND
314 314 315 FT TURN
316 316 316 FT STRAND
320 320 321 FT TURN
324 324 327 FT STRAND
331 331 335 FT STRAND
336 336 338 FT TURN
339 339 343 FT STRAND
346 346 362 FT HELIX
363 363 364 FT TURN
367 367 376 FT HELIX
377 377 377 FT TURN
379 379 379 FT TURN
382 382 383 FT TURN
384 384 384 FT STRAND
385 385 387 FT TURN
388 388 388 FT STRAND
391 391 392 FT STRAND
395 395 399 FT HELIX
400 400 401 FT TURN
SQ SEQUENCE 401 AA; 42835 MW; 1C736EFAA9F256F CRC64;

Query Match 10.8%; Score 242.5; DB 1; Length 401;
Best Local Similarity 30.9%; Pred. No. 8e-09;
Matches 81; Conservative 30; Mismatches 100; Indels 51; Gaps 10;

QY 11 DVAGSSYGLYGGQIVAVADTGLDTGNDSSMHEAPRGKITALLY-ALGRTNNANDTNGHC 69
DB 146 DVTKGS-----SQGTAVDITGVD-----YTHPDLGKVIKGYDFVNDYDPMDLNNHG 194
QY 70 THVAG---SVLGNSTNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYSAGARIH 126
DB 195 THVAGIAAETNNATGIAGVAPNTRILAVRALDRNG--SGTSLSDIADIAYAADSGAEVI 252
QY 127 TNSWGAANGAYTTDSRVDYVRKNDMTILPAAGNEGNGGTISAPGTAKNAITVGATE 186
DB 253 NLSLGC---DCHTTLENVAVNYAWNKGSVVVAAAGNNG--SSTTFPPASVENVIAVGA-- 305
QY 187 NLRPSFGSYADNHNHVAQFSRGTGDKRIKPDVWAPGTFILSARSSLAPDSSFWANHDS 246
DB 306 -----VDQYDLASFSNYGTW-----VDVWAPGVDIVSTITG-----N 338

```



```

Db 166 ANAMAVGATDQNNRASFQYAGI-----DIVAPGVNVQSTYPG- 205
QY 235 APDSSFANHDSKYAYMGTSVATPIVAGNVA-----QLREHFVKN 275
Db 206 -----STYASLNGTSMATPHVAGAAALVKQNPWSNVQIRNH-LKN 246

RESULT 13
ELYA_BACAO
ID ELYA_BACAO STANDARD; PRT; 380 AA.
AC P27693;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alkaline protease precursor (EC 3.4.21.-).
OS Bacillus alcalophilus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1445;
RN [1]
RC STRAIN=PB92;
RX MEDLINE=91282483; PubMed=2059048;
RA van der Laan J.C., Gerritse G., Mulleners L.J.M., van der Hoek R.A.,
RA Quax W.J.;
RT "Cloning, characterization, and multiple chromosomal integration of a
RT Bacillus alkaline protease gene.";
RL Appl. Environ. Microbiol. 57:901-909(1991).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.75 ANGSTROMS).
RC STRAIN=PB92;
RX MEDLINE=92390330; PubMed=1518788;
RA van der Laan J.C., Teplyakov A.V., Kelders H., Kaik K.H., Misset O.,
RA Mulleners L.J.M., Dijkstra B.W.;
RT "Crystal structure of the high-alkaline serine protease PB92 from
RT Bacillus alcalophilus.";
RL Protein Eng. 5:405-411(1992).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS).
RX MEDLINE=93078250; PubMed=1447775;
RA Sobek H., Hecht H.-J., Aehle W., Schomburg D.;
RT "X-ray structure determination and comparison of two crystal forms of
RT a variant (Asn118Arg) of the alkaline protease from Bacillus
RT alcalophilus refined at 1.85-A resolution.";
RL J. Mol. Biol. 228:108-117(1992).
RN [4]
RP STRUCTURE BY NMR OF 112-380.
RC STRAIN=PB92;
RX Martin J.R., Mulder F.A., Karimi-Nejad Y., van der Zwan J.,
RA Mariani M., Schipper D., Boelens R.;
RT "The solution structure of serine protease PB92 from Bacillus
RT alcalophilus presents a rigid fold with a flexible substrate-binding
RT site.";
RL Structure 5:521-532(1997).
CC -!- COFACTOR: Binds 2 calcium ions per subunit.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to peptidase family S8.

-----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M65086; AAA22212.1; -
DR EMBL; A13738; CAA01128.1; -
DR PIR; A49778; A49778.
DR PDB; 1AH2; 15-APR-98.
DR MEROPS; S08.038; -.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR009020; Protease_inhib.

```

```

DR Pfam; PF00082; Peptidase_S8; 1.
DR PRIN2; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
KW Hydrolase; Serine protease; Zymogen; Metal-binding; Calcium-binding;
KW Signal; 3D-structure.
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 112
FT CHAIN 113 380 ALKALINE PROTEASE.
FT ACT_SITE 143 143 CHARGE RELAY SYSTEM.
FT ACT_SITE 173 173 CHARGE RELAY SYSTEM.
FT ACT_SITE 326 326 CHARGE RELAY SYSTEM.
FT METAL 113 113 CALCIUM 1.
FT METAL 151 151 CALCIUM 1.
FT METAL 184 184 CALCIUM 1 (VIA CARBONYL OXYGEN).
FT METAL 186 186 CALCIUM 1.
FT METAL 188 188 CALCIUM 1 (VIA CARBONYL OXYGEN).
FT METAL 190 190 CALCIUM 1 (VIA CARBONYL OXYGEN).
FT METAL 274 274 CALCIUM 2 (VIA CARBONYL OXYGEN).
FT METAL 276 276 CALCIUM 2 (VIA CARBONYL OXYGEN).
FT METAL 279 279 CALCIUM 2 (VIA CARBONYL OXYGEN).
FT HELIX 117 120
FT TURN 121 123
FT HELIX 125 130
FT TURN 131 131
FT TURN 135 136
FT STRAND 138 142
FT TURN 150 151
FT STRAND 156 157
FT HELIX 173 182
FT TURN 197 201
FT STRAND 207 208
FT HELIX 213 226
FT STRAND 230 233
FT HELIX 242 252
FT TURN 253 255
FT STRAND 257 258
FT TURN 267 269
FT TURN 272 275
FT STRAND 279 279
FT STRAND 282 285
FT STRAND 291 291
FT TURN 299 300
FT STRAND 303 306
FT STRAND 310 314
FT TURN 315 317
FT STRAND 318 322
FT HELIX 325 342
FT TURN 344 345
FT HELIX 348 358
FT HELIX 375 380
SQ SEQUENCE 380 AA; 539EA72771B6682C CRC64;
Query Match 10.7%; Score 240; DB 1; Length 380;
Best Local Similarity 31.1%; Pred. No. 1.1e-08;
Matches 91; Conservative 30; Mismatches 90; Indels 82; Gaps 14;

```

```

QY 8 VKADVAGSSVGLYQCGQIVAVADTGLDTCRNDSSMHEAFRGKITALYALGRTRNANDTNG 67
Db 122 VQAPAAHNR-GLTSGGVKVAVLDTGIST-----HPDLNIRGGASFVPGEP-STODGNG 172
QY 68 HGTHVAGSV--LNGSGSTNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFQAYSAGARI 125
Db 173 HGTHVAGTIAALNNSIGVLGVAFNAELYAVKVLGASG--SGSVSSTAQGLEWAGNNGMHV 230
QY 126 HTNSWGA-----AVNGAYTDSRNVDDYVRKNDMTILFAAGNEGNGGTISAPGTA 176
Db 231 ANLSLGSPPSATLEQAVNSA---TSRGV-----LVVAASGNSG--AGSISYPARY 276
QY 177 KNATVGAETE--NLRPSFGSYADNINHVAFSSRGPTKGRIPKPDVMAFGTILSARSSL 234
Db 277 ANAVAVGATDQNNRASFQYAGI-----DIVAPGVNVQSTYPG- 316

```

QY 235 APDSFWANHDSKYAYMGTSMTPIVAGNVA-----OLREHFVN 275
 Db 317 -----STYASLNGTSMATPHVAGAAALVKQKNSWSNVQIRNH-LKN 357

RESULT 14
 ID ELVA_BACCS STANDARD; PRT; 380 AA.
 AC P41362;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Alkaline protease precursor (EC 3.4.21.-)
 OS Bacillus clausii.
 OS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=9880;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;
 RX MEDLINE=93043753; PubMed=1368952;
 RA Takami H., Kobayashi T., Kobayashi M., Yamamoto M., Nakamura S.,
 Aono R., Horikoshi K.;
 RT "Molecular cloning, nucleotide sequence, and expression of the
 RT structural gene for alkaline serine protease from alkaliphilic
 RT Bacillus sp. 221.";
 RL Biosci. Biotechnol. Biochem. 56:1455-1460(1992).
 [2]
 RN SEQUENCE OF 112-129.
 RC STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;
 RA Horikoshi K.;
 RL (In) Horikoshi K. (eds.);
 RL Microorganisms in alkaline environments, pp.187-194, VCH,
 RL Weinheim (1991).
 CC -!- COFACTOR: Binds 2 calcium ions per subunit (By similarity).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; S48754; AAC60420.1; -;
 CC EMBL; D13157; BAA02442.1; -;
 CC EMBL; A26817; CAA01836.1; -;
 CC EMBL; A22550; CAA01611.1; -;
 CC HSP; P29600; LC1.
 CC MEROPS; S08.103; -;
 CC InterPro; IPR000209; Peptidase_S8.
 CC InterPro; IPR009020; Protease_inhib.
 CC Pfam; PF00082; Peptidase_S8; 1.
 CC PRINTS; PR00723; SUBTILISIN.
 CC PROSITE; PS00136; SUBTILASE_HIS; 1.
 CC PROSITE; PS00137; SUBTILASE_HIS; 1.
 CC PROSITE; PS00138; SUBTILASE_SER; 1.
 CC Hydrolase; Serine protease; Zymogen; Metal-binding; Calcium-binding;
 CC Signal.
 KW SIGNAL.
 FT SIGNAL 1 27 POTENTIAL.
 FT PROPEP 28 111
 FT CHAIN 112 380
 FT ACT_SITE 143 143
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 173 173
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 326 326
 FT ACT_SITE 326 326
 FT METAL 113 113
 FT METAL 151 151
 FT METAL 184 184
 FT METAL 184 184
 FT METAL 186 186
 FT METAL 186 186
 FT METAL 188 188

FT METAL 190 190
 FT METAL 274 274
 FT METAL 274 274
 FT METAL 276 276
 FT METAL 276 276
 FT METAL 279 279
 FT METAL 279 279
 SQ SEQUENCE 380 AA; 38826 MW; 5F73ABC68D5B6831 CRC64;
 Query Match 10.7%; Score 240; DB 1; Length 380;
 Best Local Similarity 31.1%; Pred. No. 1.1e-08;
 Matches 91; Conservative 30; Mismatches 90; Indels 82; Gaps 14;
 QY 8 VKADVAOSSYGLYGGQIVAVADTGLDTGRNDSMHEAFRGKITALYALGRITNNANTNG 67
 Db 122 VQAPAHNR-GLTSGYKVAVLDTGIST-----HPDLNIRGGASFVPGEP-STQDGG 172
 QY 68 HGTHTVAGSV--LNGSTNKGMAPOANLVQISIMDSGGGLGSLPSNLQTLFSQAYSAGARI 125
 Db 173 HGTHVAGTIAALNNSIGVLGVAPSAELYAVKVLGASG--SGSVSSIAQGLEWAGNNGMHV 230
 QY 126 HTNSWGA-----AVNGAYTTDSRNVDYVRKNDMTILFAAGNENGNGGTISAPGTA 176
 Db 231 ANLSGSPSPSATLEQAVNSA---TSRGV-----LVVAASGNSG--AGSISYPARY 276
 QY 177 KNAITVGATE--NLRPSPGSYADNINHVAFSSRGPTKQGRKIPDVMAPGTFIISARSSL 234
 Db 277 ANAMAVGATDQNNNRASPSQYAGL-----DIVAPGVNVQSTYPG- 316
 QY 235 APDSFWANHDSKYAYMGTSMTPIVAGNVA-----OLREHFVN 275
 Db 317 -----STYASLNGTSMATPHVAGAAALVKQKNSWSNVQIRNH-LKN 357

RESULT 15
 ID ELVA_BACCS STANDARD; PRT; 269 AA.
 AC Q99405;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE M-protease (EC 3.4.21.-)
 OS Bacillus sp. (strain KSM-K16).
 OS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1409;
 RN [1]
 RN X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
 RA Yamane T., Kani T., Hatanaka T., Suzuki A., Ashida T.,
 RA Kobayashi T., Ito S., Yamashita O.;
 RT "Structure of a new alkaline serine protease (M-protease) from
 RT Bacillus sp. KSM-K16.";
 RL Acta Crystallogr. D 51:199-206(1995).
 RN [2]
 RN SEQUENCE OF 1-23, AND CHARACTERIZATION.
 RX MEDLINE=95358832; PubMed=7632397;
 RA Kobayashi T., Hakamada Y., Adachi S., Hitomi J., Yoshimatsu T.,
 RA Koike K., Kawai S., Ito S.;
 RT "Purification and properties of an alkaline protease from
 RT alkaliphilic Bacillus sp. KSM-K16.";
 RL Appl. Microbiol. Biotechnol. 43:473-481(1995).
 CC -!- COFACTOR: Binds 2 calcium ions per subunit.
 CC -!- SUBUNIT: Monomer.
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC PDB; 1MPT; 22-JUN-94.
 DR MEROPS; S08.010; -;
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE_HIS; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.

KW Hydrolase; Serine protease; Metal-binding; Calcium-binding;
3D-structure. 32 32 CHARGE RELAY SYSTEM.
FT ACT SITE 62 62 CHARGE RELAY SYSTEM.
FT ACT SITE 215 215 CHARGE RELAY SYSTEM.
FT METAL 2 2 CALCIUM 1.
FT METAL 40 40 CALCIUM 1.
FT METAL 73 73 CALCIUM 1 (VIA CARBONYL OXYGEN).
FT METAL 75 75 CALCIUM 1.
FT METAL 77 77 CALCIUM 1 (VIA CARBONYL OXYGEN).
FT METAL 79 79 CALCIUM 1 (VIA CARBONYL OXYGEN).
FT METAL 163 163 CALCIUM 2 (VIA CARBONYL OXYGEN).
FT METAL 165 165 CALCIUM 2 (VIA CARBONYL OXYGEN).
FT METAL 168 168 CALCIUM 2 (VIA CARBONYL OXYGEN).
FT HELIX 6 10
FT TURN 11 12
FT HELIX 13 18
FT TURN 19 20
FT TURN 24 25
FT STRAND 27 32
FT TURN 39 40
FT STRAND 43 48
FT TURN 51 52
FT HELIX 62 71
FT TURN 84 85
FT STRAND 87 92
FT TURN 96 97
FT HELIX 102 114
FT TURN 115 116
FT STRAND 119 122
FT STRAND 126 126
FT HELIX 131 142
FT TURN 143 144
FT STRAND 146 150
FT STRAND 161 161
FT TURN 162 164
FT TURN 166 167
FT STRAND 169 174
FT TURN 176 177
FT STRAND 180 180
FT TURN 182 183
FT TURN 188 189
FT STRAND 192 195
FT STRAND 199 203
FT TURN 204 206
FT STRAND 207 211
FT HELIX 214 231
FT TURN 233 234
FT HELIX 237 246
FT TURN 247 247
FT STRAND 249 249
FT HELIX 254 257
FT TURN 258 259
FT STRAND 261 261
FT HELIX 264 267
FT TURN 268 269
SQ SEQUENCE 269 AA; 26723 MW; 7A03C86D534A1D07 CRC64;

Query Match 10.6%; Score 239; DB 1; Length 269;
Best Local Similarity 31.1%; Pred. No. 8e-09;
Matches 9; Conservative 30; Mismatches 90; Indels 92; Gaps 14;

Qy 8 VKADVAQSSYGLYCGQGIIVADTGLDGTGRNDSMEHAFRGKITALYALGRTNNDNTNG 67
Db 11 VQAPAAHNR-GLAGSGVKVAVLDTGIST-----HPDLNIRGGASFVPEP-STQDQNG 61
Qy 68 HGTHVAGSV--LGNSTNKGMAQANLVFQSIINDSGGLGILPSNLQTLFSCAYSAGARI 125
Db 62 HGTHVAGTIAALNNISGLVAFPSAELYAVKVLGASG--SGSVSSIAQGLEWAGNNGMHV 119
Qy 126 HTNSWGA-----AVNGAYTTDSRNVDYVRKNDWILFAAGNEGPNGGTISAPGTA 176
Db 120 ANLSLGSFSPSATILEQAVNSA---TSRGV-----LVVAASGNSG--AGSISYPARY 165

Qy 177 KNAITVGATE--NLRPFGSYADNINHVAQFSSRGPTKGRIKPDVMAPTILSARSSL 234
Db 166 ANAMAVGATDQNNNRASFQYAGL-----DIVAFGVNVQSTYPG- 205
Qy 235 APDSFMANHDSKYAYMGGTSMATPIVAGNVA-----QUREHFVN 275
Db 206 -----STYASLNGTSMATPHVAGVAALVKQKPNFSWSNVQIRNH-LKN 246

Search completed: March 10, 2004, 14:45:18
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:42:22 ; Search time 46 Seconds
(without alignments)
2976.846 Million cell updates/sec

Title: US-09-985-689a-1
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAYNVPGPQTFSLAIVN 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 25:*

- 1: sp archaea:*
- 2: sp bacteria:*
- 3: sp fungi:*
- 4: sp human:*
- 5: sp invertebrate:*
- 6: sp mammal:*
- 7: sp mhc:*
- 8: sp organelle:*
- 9: sp phage:*
- 10: sp plant:*
- 11: sp rodent:*
- 12: sp virus:*
- 13: sp vertebrate:*
- 14: sp unclassified:*
- 15: sp virus:*
- 16: sp bacteriap:*
- 17: sp archaep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2247	100.0	640	2 Q93UV9	Q93UV9 bacillus sp
2	2183	97.2	639	2 Q9AQR3	Q9AQR3 bacillus sp
3	2143	95.4	434	2 Q9AQR0	Q9AQR0 bacillus sp
4	1998.5	88.9	433	2 Q9AQR1	Q9AQR1 bacillus sp
5	1994.5	88.8	433	2 Q9AQR4	Q9AQR4 bacillus sp
6	1987.5	88.5	433	2 Q9AQR2	Q9AQR2 bacillus sp
7	511.5	22.8	5	Q8T9W1	Q8T9W1 dictyosteli
8	447	19.9	1702	5 Q8T9W1	Q8T9W1 dictyosteli
9	414	18.4	654	17 Q8UOC9	Q8UOC9 pyrococcus
10	406.5	18.1	561	16 Q8RBZ2	Q8RBZ2 thermoanaer
11	398	17.7	1239	16 Q9FBZ4	Q9FBZ4 streptomyc
12	381	17.0	430	16 Q8ENV1	Q8ENV1 oceanobacil
13	376	16.7	1253	16 Q9FCO6	Q9FCO6 streptomyc
14	366	16.3	1102	2 P95684	P95684 streptomyc
15	363.5	16.2	1208	16 Q82B14	Q82B14 streptomyc
16	349	15.5	1245	16 Q9RL54	Q9RL54 streptomyc

17	347.5	15.5	444	16	Q9KB77	Q9KB77 bacillus ha
18	346.5	15.4	1139	16	Q82I39	Q82I39 streptomyc
19	346	15.4	1105	2	Q8KKH6	Q8KKH6 streptomyc
20	338	15.0	412	2	Q9AER6	Q9AER6 thermoanaer
21	338	15.0	1237	2	Q8GTT4	Q8GTT4 streptomyc
22	337	15.0	412	16	Q8RC68	Q8RC68 thermoanaer
23	328.5	14.6	1220	16	Q8L0A0	Q8L0A0 streptomyc
24	327.5	14.6	435	16	Q8EMJ3	Q8EMJ3 oceanobacil
25	324.5	14.4	824	2	Q45464	Q45464 bacillus su
26	320.5	14.3	442	16	Q31788	Q31788 bacillus su
27	320.5	14.3	891	1	Q93635	Q93635 thermococu
28	308.5	13.7	431	2	Q933L6	Q933L6 bacillus sp
29	306.5	13.6	1398	1	Q9P9L1	Q9P9L1 pyrococcus
30	297	13.2	1135	1	Q9P9D1	Q9P9D1 uncultured
31	294.5	13.1	434	2	Q54327	Q54327 bacillus sp
32	282.5	12.6	586	16	Q8PAL8	Q8PAL8 xanthomonas
33	281	12.5	419	2	Q45681	Q45681 bacillus su
34	280	12.5	799	16	Q9KEM1	Q9KEM1 bacillus ha
35	279.5	12.4	1345	1	Q54437	Q54437 staphylothe
36	278.5	12.4	959	16	Q8PKS7	Q8PKS7 xanthomonas
37	277.5	12.3	621	2	Q9F486	Q9F486 alteromonas
38	277.5	12.3	621	2	Q53401	Q53401 alteromonas
39	272.5	12.1	568	16	Q8PMCO	Q8PMCO xanthomonas
40	272.5	12.1	575	16	Q8PNW1	Q8PNW1 xanthomonas
41	270	12.0	1101	16	Q82CF0	Q82CF0 streptomyc
42	268.5	11.9	403	2	Q45463	Q45463 bacillus sp
43	268.5	11.9	814	16	Q82VB3	Q82VB3 nitrosomona
44	267	11.9	715	2	P70765	P70765 alteromonas
45	265	11.8	1098	16	Q9L128	Q9L128 streptomyc

ALIGNMENTS

RESULT 1

Q93UV9 PRELIMINARY; PRT; 640 AA.
AC Q93UV9; TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 24, Last annotation update)
DE Protease.
GN PROF.
OS Bacillus sp. KSM-KP43.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=109322;
RN [1]_TaxID=109322;
RP SEQUENCE FROM N.A.
RC STRAIN=KP43;
RA Itoh S., Saeki K.;
RT "new protease."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB051423; BAB55674.2; -
DR GO; GO:0004289; P:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
SQ SEQUENCE 640 AA; 67991 MW; 48BAF77E9D592C15 CRC64;

Query Match 100.0%; Score 2247; DB 2; Length 640;
Best Local Similarity 100.0%; Pred. No. 3.8e-122;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NDVARGIVKADVAQSSYGLYGGQGVAVADTGLDTGRNDSSMHEAFRGKITALVALGRIN 60
Db 207 NDVARGIVKADVAQSSYGLYGGQGVAVADTGLDTGRNDSSMHEAFRGKITALVALGRIN 266
Qy 61 NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFOSIMDSGGGLGSLPNTLQTLFSQAYS 120


```

Db      267  NNDNTNGHGTAVAGSVLNGSTNKGMAPOANLVFQSIMDSSGGGLPSNLQTLFSQAYS 326
QY      121  AGARIHNTNSGAAVNGAYTTDSRNVDYVVKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db      327  AGARIHNTNSGAAVNGAYTTDSRNVDYVVKNDMTILFAAGNEGPNGGTISAPGTAKNAI 386
QY      181  TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIPKDPVMAFGTFILSARSSLPDSSF 240
Db      387  TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIPKDPVMAFGTFILSARSSLPDSSF 446
QY      241  WANHDSKYAYMGTSWATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAAGADIGLY 300
Db      447  WANHDSKYAYMGTSWATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAAGADIGLY 506
QY      301  PNGNQGWGRVTLDKSLNVAYNVNESSLSSTQKATYSFTATAGKPLKISLWSDAPASTTA 360
Db      507  PNGNQGWGRVTLDKSLNVAYNVNESSLSSTQKATYSFTATAGKPLKISLWSDAPASTTA 566
QY      361  SVTLVNDLVLITAPNGTQVGNDFTSFYNDWNGRNNVNFVINAQSGTGTIEVQAYN 420
Db      567  SVTLVNDLVLITAPNGTQVGNDFTSFYNDWNGRNNVNFVINAQSGTGTIEVQAYN 626
QY      421  VPVGPQTFSLAIVN 434
Db      627  VPVGPQTFSLAIVN 640

```

RESULT 2

```

Q9AQR3 ID Q9AQR3 PRELIMINARY; PRT; 639 AA.
AC Q9AQR3;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Protease.
GN PROA.
OS Bacillus sp. 9860.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=133778;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9860;
RX MEDLINE=20568675; PubMed=1118284;
RA Saeiki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus sp.: enzymatic properties, sequences, and
RT evolutionary relationships."
RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
DR EMBL; AB046403; BAB21266.2; -.
DR HSSP; P00782; 1SUP.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR Pfam; PF04151; PPC; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
SQ SEQUENCE 639 AA; 68185 MW; 316AF6FDBE4FF54 CRC64;

```

Query Match 97.3%; Score 2183; DB 2; Length 639;
 Best Local Similarity 96.3%; Pred. No. 1.9e-118;
 Matches 418; Conservative 13; Mismatches 3; Indels 0; Gaps 0;

```

QY      1  NDVARGIVKADVQAQSSVGLYGQGIYVADTGLDGRNDSMHEAPRGKITALYALGRTN 60
Db      206  NDVARGIVKADVQAQSSVGLYGQGIYVADTGLDGRNDSMHEAPRGKITALYALGRTN 265
QY      61  NNDNTNGHGTAVAGSVLNGSTNKGMAPOANLVFQSIMDSSGGGLPSNLQTLFSQAYS 120

```

```

Db      266  NNDNTNGHGTAVAGSVLNGSTNKGMAPOANLVFQSIMDSSGGGLPSNLQTLFSQAFS 325
QY      121  AGARIHNTNSGAAVNGAYTTDSRNVDYVVKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db      326  AGARIHNTNSGAAVNGAYTTDSRNVDYVVKNDMTILFAAGNEGPNGGTISAPGTAKNAI 385
QY      181  TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIPKDPVMAFGTFILSARSSLPDSSF 240
Db      386  TVGATENLRPSFGSYADNINHVAQFSRSGPTKDGRIPKDPVMAFGTFILSARSSLPDSSF 445
QY      241  WANHDSKYAYMGTSWATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAAGADIGLY 300
Db      446  WANHDSKYAYMGTSWATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAAGADIGLY 505
QY      301  PNGNQGWGRVTLDKSLNVAYNVNESSLSSTQKATYSFTATAGKPLKISLWSDAPASTTA 360
Db      506  PNGNQGWGRVTLDKSLNVAYNVNESSLSSTQKATYSFTATAGKPLKISLWSDAPASTTA 565
QY      361  SVTLVNDLVLITAPNGTQVGNDFTSFYNDWNGRNNVNFVINAQSGTGTIEVQAYN 420
Db      566  SVTLVNDLVLITAPNGTQVGNDFTSFYNDWNGRNNVNFVINAQSGTGTIEVQAYN 625
QY      421  VPVGPQTFSLAIVN 434
Db      626  VPVGPQTFSLAIVN 639

```

RESULT 3

```

Q9AQR0 ID Q9AQR0 PRELIMINARY; PRT; 434 AA.
AC Q9AQR0;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Protease (Fragment).
GN PROE.
OS Bacillus sp. NV1.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=133781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NV1;
RX MEDLINE=20568675; PubMed=1118284;
RA Saeiki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus sp.: enzymatic properties, sequences, and
RT evolutionary relationships."
RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY 98.
DR EMBL; AB046406; BAB21269.1; -.
DR HSSP; P00782; 1SUP.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR Pfam; PF04151; PPC; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydroxylase; Protease; Serine protease.
FT NON_TER 1
FT NON_TER 434
SQ SEQUENCE 434 AA; 45294 MW; 83517EDDB74125D2 CRC64;

```

Query Match 95.4%; Score 2143; DB 2; Length 434;
 Best Local Similarity 93.5%; Pred. No. 2.4e-116;
 Matches 406; Conservative 19; Mismatches 9; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVQAQSSVGLYGQGIYVADTGLDGRNDSMHEAPRGKITALYALGRTN 60

Db 1 NDVARGIKADVAQSSVGLYGGQVAVADTGLDTGRNDSSMHEAFRGKITAIYALGRTN 60
QY 61 NNDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSGIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 61 NNDPNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSVMDSSNGGLGGLPSNVSTLFSQAYS 120
QY 121 AGARIHTNSGAAVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db 121 AGARIHTNSGAPVNGAYTTDSNRVDDYVRKNDMAVLFAGNEGPNNGGTISAPGTAKNAI 180
QY 181 TVGATENLRPSFGSYADININHAQFSSRGPTKGRIPKDPVMACTFILSARSSLAPDSF 240
Db 181 TVGATENLRPSFGSYADININHAQFSSRGPTKGRIPKDPVMACTFILSARSSLAPDSF 240
QY 241 WANHSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLGY 300
Db 241 WANHSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLGY 300
QY 301 PNGNGGGRVTLDSKSLNVAVYVNESSTLSOKATYSFTATAGKPLKISLWSDAPASTTA 360
Db 301 PNGNGGGRVTLDSKSLNVAVYVNESSTLSOKATYSFTATAGKPLKISLWSDAPASTTA 360
QY 361 SVTLVNDLDLVTAPNGTQVGNDFTSFYNDNWDGRNNVFNAPQSGTYTIEVQAYN 420
Db 361 SVTLVNDLDLVTAPNGTQVGNDFTSFYNDNWDGRNNVFNAPQSGTYTIEVQAYN 420
QY 421 VPVGPQTFSLAIVN 434
Db 421 VPVGPQTFSLAIVN 434

RESULT 4
Q9AQR1 ID Q9AQR1 PRELIMINARY; PRT; 433 AA.
AC Q9AQR1
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Protease (Fragment).
GN PROD.
OS Bacillus sp. SD521.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=133780;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=SD521.
RX MEDLINE=20568675; PubMed=11118284;
RA Saeki K., Okuda Y., Hatada Y., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.;
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
RT evolutionary relationships.";
RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
DR EMBL; AB046405; BAB2128.1; -;
DR HSSP; Q45670; 1DBI.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF04151; PPC; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
KW Hydrolase; Protease; Serine protease.
FT NON_TER 1
FT NON_TER 433
SQ SEQUENCE 433 AA; 45576 MW; 98A2DF18FE65DDC CRC64;
Query Match 88.9%; Score 1998.5; DB 2; Length 433;
Best Local Similarity 87.8%; Pred. No. 5.5e-108;

Matches 381; Conservative 28; Mismatches 24; Indels 1; Gaps 1;
QY 1 NDVARGIKADVAQSSVGLYGGQVAVADTGLDTGRNDSSMHEAFRGKITAIYALGRTN 60
Db 1 NDVARGIKADVAQSSVGLYGGQVAVADTGLDTGRNDSSMHEAFRGKITAIYALGRTN 60
QY 61 NNDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSGIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 61 NNDPNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSVMDSSNGGLGGLPSNVSTLFSQAYS 120
QY 121 AGARIHTNSGAAVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db 121 AGARIHTNSGAPVNGAYTTDSNRVDDYVRKNDMAVLFAGNEGPNNGGTISAPGTAKNAI 180
QY 181 TVGATENLRPSFGSYADININHAQFSSRGPTKGRIPKDPVMACTFILSARSSLAPDSF 240
Db 181 TVGATENLRPSFGSYADININHAQFSSRGPTKGRIPKDPVMACTFILSARSSLAPDSF 240
QY 241 WANHSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLGY 300
Db 241 WANHSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLGY 300
QY 301 PNGNGGGRVTLDSKSLNVAVYVNESSTLSOKATYSFTATAGKPLKISLWSDAPASTTA 360
Db 301 PNGNGGGRVTLDSKSLNVAVYVNESSTLSOKATYSFTATAGKPLKISLWSDAPASTTA 360
QY 361 SVTLVNDLDLVTAPNGTQVGNDFTSFYNDNWDGRNNVFNAPQSGTYTIEVQAYN 420
Db 361 SVTLVNDLDLVTAPNGTQVGNDFTSFYNDNWDGRNNVFNAPQSGTYTIEVQAYN 420
QY 421 VPVGPQTFSLAIVN 434
Db 421 VPVGPQTFSLAIVN 434
RESULT 5
Q9AQR4 ID Q9AQR4 PRELIMINARY; PRT; 433 AA.
AC Q9AQR4
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Protease (Fragment).
GN PROD.
OS Bacillus sp. D6.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=127889;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=D6;
RX MEDLINE=20568675; PubMed=11118284;
RA Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.;
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
RT evolutionary relationships.";
RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
DR EMBL; AB046402; BAB21265.1; -;
DR HSSP; Q45670; 1DBI.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF04151; PPC; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
KW Hydrolase; Protease; Serine protease.
FT NON_TER 1
FT NON_TER 433

SQ SEQUENCE 433 AA; 45636 MW; 52087BOA2516107F CRC64;

Query Match 88.8%; Score 1994.5; DB 2; Length 433;
Best Local Similarity 87.6%; Pred. No. 9.4e-108;
Matches 380; Conservative 28; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDGRNDSSMHEAFRGKITALLYALGRTN 60
DB 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDGRNDSSMHEAFRGKITALLYALGRTN 60

QY 61 NANDTNGHGHVAGSVLGN-ALNKGMAPQANLVFQSIMDSGGGLGGLPSNLTLFSQAWN 120
DB 61 NANDTNGHGHVAGSVLGN-ALNKGMAPQANLVFQSIMDSGGGLGGLPSNLTLFSQAWN 119

QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILPAAGNEGPNGGTISAPGTAKNAI 180
DB 120 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILPAAGNEGPNGGTISAPGTAKNAI 179

QY 181 TVGATENLRPSFGSIADNPNHIAQFSSRGATRDGRIKPDVTAFTILSARSSLPDSSF 240
DB 180 TVGATENLRPSFGSIADNPNHIAQFSSRGATRDGRIKPDVTAFTILSARSSLPDSSF 239

QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGADIGLY 300
DB 240 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGADIGLY 299

QY 301 PNGNQGWRVTLKSLNVAVNEATLTGQKATYSFQACGKPLKISLWSDAPASTTA 360
DB 300 PNGNQGWRVTLKSLNVAVNEATLTGQKATYSFQACGKPLKISLWSDAPASTTA 359

QY 361 SVTLVNDLVLITAPNGTQYVGNDFSPYNDNWDGNNVFNAPQSGTYYIEVOAYN 420
DB 360 SVTLVNDLVLITAPNGTQYVGNDFSPYNDNWDGNNVFNAPQSGTYYIEVOAYN 419

QY 421 VPVGQPTFSLAIYN 434
DB 420 VPVGQPTFSLAIYH 433

RESULT 6

Q9AQR2 ID Q9AQR2 PRELIMINARY; PRT; 433 AA.

AC Q9AQR2;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Protease (Fragment).

GN PROC.

OS Bacillus sp. Y.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

CX NCBI_TaxID=133779;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Y;

RX MEDLINE=20568675; PubMed=11118284;

RA Saeki K.; Okuda M.; Hatada Y.; Kobayashi T.; Ito S.; Takami H.;

RA Horikoshi K.;

RT "Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic Bacillus spp.: enzymatic properties, sequences, and evolutionary relationships.";

RL Biochem. Biophys. Res. Commun. 279:313-319(2000).

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.

DR EMBL; AB046404; BAB21267.1; -.

DR HSP; Q45670; 1DB1.

DR GO; GO:0008233; F:peptidase activity; IEA.

DR GO; GO:0004289; F:subtilase activity; IEA.

DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

DR InterPro; IPR000209; Peptidase_S8.

DR Pfam; PF00082; Peptidase_S8; 1.

DR Pfam; PF04151; PPC; 1.

DR PRINTS; PR00723; SUBTILISIN.

DR PROSITE; PS00137; SUBTILASE_HIS; 1.

DR PROSITE; PS00138; SUBTILASE_SER; 1.

KW Hydrolase; Protease; Serine protease.

FT NON_TER 433 1

SQ SEQUENCE 433 AA; 45587 MW; B81291A803C75AE CRC64;

Query Match 88.5%; Score 1987.5; DB 2; Length 433;
Best Local Similarity 87.3%; Pred. No. 2.4e-107;
Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDGRNDSSMHEAFRGKITALLYALGRTN 60
DB 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDGRNDSSMHEAFRGKITALLYALGRTN 60

QY 61 NANDTNGHGHVAGSVLGN-ALNKGMAPQANLVFQSIMDSGGGLGGLPSNLTLFSQAWN 120
DB 61 NANDTNGHGHVAGSVLGN-ALNKGMAPQANLVFQSIMDSGGGLGGLPSNLTLFSQAWN 119

QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILPAAGNEGPNGGTISAPGTAKNAI 180
DB 120 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILPAAGNEGPNGGTISAPGTAKNAI 179

QY 181 TVGATENLRPSFGSIADNPNHIAQFSSRGATRDGRIKPDVTAFTILSARSSLPDSSF 240
DB 180 TVGATENLRPSFGSIADNPNHIAQFSSRGATRDGRIKPDVTAFTILSARSSLPDSSF 239

QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGADIGLY 300
DB 240 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGADIGLY 299

QY 301 PNGNQGWRVTLKSLNVAVNEATLTGQKATYSFQACGKPLKISLWSDAPASTTA 360
DB 300 PNGNQGWRVTLKSLNVAVNEATLTGQKATYSFQACGKPLKISLWSDAPASTTA 359

QY 361 SVTLVNDLVLITAPNGTQYVGNDFSPYNDNWDGNNVFNAPQSGTYYIEVOAYN 420
DB 360 SVTLVNDLVLITAPNGTQYVGNDFSPYNDNWDGNNVFNAPQSGTYYIEVOAYN 419

QY 421 VPVGQPTFSLAIYN 434
DB 420 VPVGQPTFSLAIYH 433

RESULT 7

Q8T9W1 ID Q8T9W1 PRELIMINARY; PRT; 1825 AA.

AC Q8T9W1;

DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Serine protease/ABC transporter TagD.

GN TAGD.

OS Dictyostelium discoideum (Slime mold).

OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.

CX NCBI_TaxID=44689;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Ax4;

RA Anjard C.; Loomis W.F.;

RT "Evolution of the ABC transporters of Dictyostelium";

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.

DR EMBL; AF466309; AAL74253.1; -.

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.

DR GO; GO:0001656; F:nucleotide binding; IEA.

DR GO; GO:0008233; F:peptidase activity; IEA.

DR GO; GO:0004289; F:subtilase activity; IEA.

DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

DR GO; GO:0006810; P:transport; IEA.

DR InterPro; IPR003593; AAA ATPase.

DR InterPro; IPR001140; ABC_TM_transp.


```
RESULT 9
Q8UBC9 PRELIMINARY; PRT; 654 AA.
AC Q8UBC9;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Alkaline serine protease.
GN PF1670.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; A8010265; AAL81794.1; -
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PRO0723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
KW Protease; Complete proteome.
SQ SEQUENCE 654 AA; 70230 MW; 1CB145A5F505DB34 CRC64;

Query Match 18.4%; Score 414; DB 17; Length 654;
Best Local Similarity 29.8%; Pred. No. 5.4e-16;
Matches 136; Conservative 59; Mismatches 147; Indels 114; Gaps 19;

QY 18 GLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRNNAN-----DTNGHGH 71
DB 154 GYDGSGITGIIDTGID-----ASHPDLOQKV-----IGWDFVNGRSYFDDHGHGH 202

QY 72 VAGSVLNGSTN-----KGMAPQNLVFGSIM--DSGGGLGLPLSNLQTLFQOAYSAGARI 125
DB 203 VASIAAGTGAASNGKYKGMAPGAKLAGIKVLGADGSGSITTIKGVAVDNKDKYGIKV 262

QY 126 HTNSMGA-----AVNGAYTDSRVDDYVRKNDMTILFAAGNEGNGGTISAP 173
DB 263 INLSGSSSGDGTALSOAVNAWDA-----GLVVVAAGNSGPNKVTIGSP 310

QY 174 GTAKNAITVATENLRPSFGSVADNINHVAFSSRGPTKDGRIKPDVMAPGTFILSARSS 233
DB 311 AAASKVITVGA-----VDKYDVITSPSSRGPTADGRKPEVAVPAGNWIILARAS 359

QY 234 LAPDSSFVANHDSKYAYMGTSNATPIVAGNVAQLREHFVKRGITPK--PSLLKALIA 291
DB 360 -----GTSMGQFINDYYTAPGTSNATPHVAGIALLLQ-----AHPSTPKVKTALIE 409

QY 292 GA-----ADIGLYPNGNGQGRVTLDKSLNVAYNNESSLSLSOKA-----TYSFT 338
DB 410 TADIVKPEIDAIYGA-----GRVNAYKAIN--YDNYAKLVFTGVYANKGSQTHQFV 460

QY 339 ATAGKPLKTLVWSDAPASTTASVTLVNDLDLVTAPNGTQVYVNDFTSPYNDNWDGRNN 396
DB 461 ISGASEVTATLYWDNAN-----SLLDLYLDPNNGNQ--VDYSYATAY-----G 501

QY 399 VENVFINAPQSGTYTIEVOAYNVVPGPQTFSIAIYN 434
DB 502 FEKGVYNTDGTWIKVYSYS---GSANYQVDVVS 534
```

RESULT 10
Q8RBJ2

```
Q8RBJ2 PRELIMINARY; PRT; 561 AA.
AC Q8RBJ2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Subtilisin-like serine proteases.
GN APRE2 OR TTE0824.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; A8013049; AM24081.1; -
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00728; PPC.
DR PRINTS; PRO0728; PPC.
DR InterPro; IPR009020; Protease_inhib.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF04151; PPC; 1.
DR PRINTS; PRO0723; SUBTILISIN.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
KW Protease; Complete proteome.
SQ SEQUENCE 561 AA; 59968 MW; BA9C5C2F7083A18 CRC64;

Query Match 18.1%; Score 406.5; DB 16; Length 561;
Best Local Similarity 31.3%; Pred. No. 1.2e-15;
Matches 144; Conservative 62; Mismatches 135; Indels 119; Gaps 22;

QY 6 GIVKADVQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRNNAN-- 63
DB 155 GITK--ARSDFGVTGKNITAIIDTGDGNHVDLS-----GGKI-----IGWDFPINK 201

QY 64 -----DTNGHGHVAGSVLNGSTN---KGMAPQNLVFGSIMDSGGG-----LGLPLSN 110
DB 202 TTPYDDNGHGHVASTAAGTGAGNSFYKGVAPDALLVIGIKVLDANGSGMSVTAGIDWA 261

QY 111 LQ-----TLFSQAYSAGARIHNSWGAANGAYTTDSRVDDYVRKNDMTILF 158
DB 262 VQNKDVVYGIKVINLSLGTSTSDG---TDSTSLAVN-----RAVD-----SGIVVVV 305

QY 159 AAGNEGNGGTISAPGTAKNAITVATENLRPSFGSVADNINHVAFSSRGPTKDGRIKP 218
DB 306 AAGNSGPAKTIISGPAEAKAITVAAMADV-----GELGFNL---ASFSSRGPTADGRIKP 358

QY 219 DVMAPGTFILSARSSLAPDSSFVANHDSKYAYMGTSNATPIVAGNVAQLREHFVKRGI 278
DB 359 DIAAPGNITAAK-----ANSVGVYVYSGTSMATPFVAGTVALMLN---ANPNL 405

QY 279 TPKPSLLKALIAAGADIGLYNGNGQGRVTLDKSLNVA-----YVNES 324
DB 406 TPNDAA--KNITNGTAKSWGPPSKNVYDYGAGRLDGYEIRVAGNFRGNNDVPHNYIT--S 461

QY 325 SSLSTSOCA--TSGTAT--AGKPLKISLV---WSDAPASTTASVTLVNDLDLVTAPNGTQ 379
DB 462 GYLFPSRYSDTWTFTNATNTSYPIATLIPDMANYNP-----DPIYLYDPSGTL 511

QY 380 YVGNDFTSFYNDNWDGRNNVFINAPQSGTYTIEVOAY 419
```


Db 173 EYDD-NGHGTGACAGNGALSQYQGPAPDANLVGVKNTGS-GSLSTVIEGI- 229
 QY 116 SQAYSAGARIHNSGAVNGAYTTDSRVDDYR-----KNDMTILFAAGNEPNGGT 169
 Db 230 DNCIQNSKINILSLSL-GSDATEPABGDPVNAVETAWDNGMVVCAAGNSGPGDKT 288
 QY 170 ISAPCTAKNAITVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAFGTFLS 229
 Db 289 VGSPIGSPKVIITVGAADNNTAERS-----DSDVAEFSRSGPTIDGLTKPNIITPGVDIVS 344
 QY 230 ABS--SLAPDSSFWANHDSKYAYMGTSMTATIVAGNVAQLREHFVKRGITPKPSLLKA 287
 Db 345 LRAPSGFIDTKNSARVNGYISLGSMTATPICAGIVAQLLQ-----SDSLT--PNQVKE 399
 QY 288 ALIAGAADIGLGPNGNGWGRVTLDKSLNVAYVNE 323
 Db 400 KLMEACQDLQGS-PN-VQAGYL---NAAANLINE 430

RESULT 13

Q9FC06 PRELIMINARY; PRT; 1253 AA.
 AC Q9FC06; MEDLINE=21996410; PubMed=12000953;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Putative secreted peptidase.
 GN SC07176 OR SC8A11.04C.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomyces; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Saunders D.C., Harris D.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Denapalte D., Eichner A., Cullum J.,
 RA Redenbach M., Kieser H.M., Hopwood D.A.;
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RP "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2)."
 RL Nature 417:141-147(2002).
 DR ENBL; A0339130; CAC01576.1; --
 DR HSSP; Q99405; IMPT.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004289; F:subtilase activity; IEA.
 DR GO; GO:000508; P:proteolysis and peptidolysis; IEA.

DR InterPro; IPR003137; PA.
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF02225; PA; 1.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Complete proteome.
 SQ SEQUENCE 1253 AA; 130971 MW; AA69B417EFEEDB89 CRC64;
 Query Match 16.7%; Score 376; DB 16; Length 1253;
 Best Local Similarity 27.8%; Pred. No. 2e-13;
 Matches 142; Conservative 58; Mismatches 145; Indels 170; Gaps 22;
 QY 16 SYGLYGQGIYAVADTGLDTGRNDSMHEAFRGKITALYALGRTNNANDTNGHGTAVGS 75
 Db 233 SGGNTGEGVAVLDTGVDAG-----HPDFAGRIATAATSFVDPQDVTDRNGHGTAVST 286
 QY 76 VLNGSTN-----KGMAPQANLVFQSIMDSGG-----GLGG 106
 Db 287 VAGTGAASGGVEKGVAPGASLHGKVLDSGSGQSDSWLWAGMEWAVRDQHAKIVMSLGD 346
 QY 107 LPSNLQTLFSQAYSAGARIHTNSWGAAGVNGAYTTDSRVDDYVRKNDMTILFAAGNEGN 166
 Db 347 SPTDGTDLSEAV-----NWLSAETGA-----LFVVAAGNSGPE 380
 QY 167 GGTISAPGTAKNAITVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAFGT 225
 Db 381 AYTGTGTPAAADAALITGVANG--PGRG-----VDQLADFSSRGPRVGDNAVKPDLTAPGV 433
 QY 226 FILSARSSLAPDSSFWANHDSKYAYMGTSMTATIVAGNVAQLREHFVKNRGITPKPSLL 285
 Db 434 GVLAARSRYAPEG-----EGAYQSLSGTSMATPHVAGAAALAAEHPDWTG-----QRL 482
 QY 286 KAALIAGAADIGLGPNGNQ-----GMGRVTLDKSLNVAYVNESSTLSQ---KATY 335
 Db 483 KEALVGTGTA-----GTQRFSPFDAGSGRV-----DVAARVSTLLASGDAFAQAHY 528
 QY 336 -----SFTATAGKPLKISLWSDA-----PASTASVTLVND 367
 Db 529 PYTPGQTVRRDVTYNSGPAPVALDLALSPAEPLPEGLFTLSEAGVTVPAHGHTASVGIITH 588
 QY 368 LD-----LVITAPNGTQYGVNDFTSPYNDNWDGNNVNFYNA-----PQSG 410
 Db 589 LAAEDNGAVATRLVAGSAGDGVLA---RTPGVNKEGRR--ATLALTAKDHDKPLSG 642
 QY 411 T-----YTIEVQA-YNVPVGPQTSL 430
 Db 643 TVILKDYERNAPKRVSVSDASGELDLRLSESTYSV 677

RESULT 14

P95684 PRELIMINARY; PRT; 1102 AA.
 ID P95684
 AC P95684; MEDLINE=97144528; PubMed=8990295;
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DE Subtilisin-like protease.
 DE Subtilisin-like protease.
 OS Streptomyces albobogiseolus.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1887;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S-3253;
 RX Suzuki M., Taguchi S., Yamada S., Kojima S., Miura K., Momose H.;
 RT "A novel member of the subtilisin-like protease family from
 RT Streptomyces albobogiseolus."
 RL J. Bacteriol. 179:430-438(1997).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 15, 2004, 22:30:03 ; Search time 3605 seconds
(without alignments)

5217.996 Million cell updates/sec

Title: US-09-985-689A-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSVGLY.....EVQAVNPVGPOTFSLAIV 434

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh

-Q=/cgn2 1/USPTO.spool/US0985689/runat_10032004_112903_19718/app query.fasta_1.583

-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0

-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45

-DOCALIGN=200 -THRESH=0.5 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL

-OUTFMT=ptco -NORM=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US0985689@cgn 1 1 2372 @runat_10032004_112903_19718 -NCPU=6 -ICPU=3

-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG

-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	2247	100.0	1305	6	AX839476	Sequence	AX839476 Sequence
2	2247	100.0	1923	1	AB051423	Bacillus	AB051423 Bacillus
3	2247	100.0	1923	6	AR368117	Sequence	AR368117 Sequence
4	2242	99.8	1923	1	AB084155	Bacillus	AB084155 Bacillus
5	2242	99.8	1923	6	AR368118	Sequence	AR368118 Sequence
6	2183	97.2	1920	1	AB046403	Bacillus	AB046403 Bacillus
7	2183	97.2	1920	6	AR368116	Sequence	AR368116 Sequence
8	2143	95.4	1302	1	AB046406	Bacillus	AB046406 Bacillus
9	2125.5	94.6	3003	6	AR069954	Sequence	AR069954 Sequence
10	2125.5	94.6	3003	6	BD062155	Nucleic a	BD062155 Nucleic a
11	1998.5	88.9	1299	1	AB046405	Bacillus	AB046405 Bacillus
12	1994.5	88.8	1299	1	AB046402	Bacillus	AB046402 Bacillus
13	1987.5	88.5	1299	1	AB046404	Bacillus	AB046404 Bacillus
14	1986.5	88.4	2218	6	E03808	DNA encodin	E03808 DNA encodin
15	632.5	28.1	6006	1	AF268611	Unculture	AF268611 Unculture
16	580	25.8	30205	1	AS017218	Geobacter	AS017218 Geobacter
17	523.5	23.3	5890	3	DDU60086	Dicystosteli	DDU60086 Dicystosteli
18	511.5	22.8	5772	3	AF466309	Dicystosteli	AF466309 Dicystosteli
19	497	22.1	6115	3	DDU20432	Dicystosteli	DDU20432 Dicystosteli
20	452.5	20.1	1977	6	AR201152	Sequence	AR201152 Sequence
21	447	19.9	5109	3	AF263455	Dicystosteli	AF263455 Dicystosteli
22	425.5	18.9	12928	2	AC096673	Trypanoso	AC096673 Trypanoso
23	416.5	18.5	1236	6	AR201146	Sequence	AR201146 Sequence
24	416.5	18.5	1962	6	AR201155	Sequence	AR201155 Sequence
25	416.5	18.5	12452	1	AB010265	Thermoana	AB010265 Thermoana
26	406.5	18.1	14376	1	AB013049	Thermoana	AB013049 Thermoana
27	398	17.7	303450	1	SC0939130	Streptomy	SC0939130 Streptomy
28	381	17.0	299850	1	AP004601	Oceanobac	AP004601 Oceanobac
29	366	16.3	3348	1	D33672	Streptomyce	D33672 Streptomyce
30	363.5	16.2	30425	1	AP005044	Streptomy	AP005044 Streptomy
31	351	15.6	2539	6	AR203322	Sequence	AR203322 Sequence
32	349	15.5	291000	1	SC0939105	Streptomy	SC0939105 Streptomy
33	347.5	15.5	292550	1	AP001513	Bacillus	AP001513 Bacillus
34	346.5	15.4	302300	1	AP005034	Streptomy	AP005034 Streptomy
35	346	15.4	2809	6	AR202321	Sequence	AR202321 Sequence
36	346	15.4	3900	1	AB007809	Streptomy	AB007809 Streptomy
37	338	15.0	1239	1	AP305633	Thermoana	AP305633 Thermoana
38	338	15.0	1239	1	AY028704	Thermoana	AY028704 Thermoana
39	338	15.0	135638	1	AF484556	Streptomy	AF484556 Streptomy
40	337	15.0	12295	1	AB013026	Thermoana	AB013026 Thermoana
41	332.5	14.8	1329	6	AX433519	Sequence	AX433519 Sequence
42	331	14.7	207829	1	BSU00010	Bacillus su	Bacillus su
43	328.5	14.6	300800	1	SC0939112	Streptomy	SC0939112 Streptomy
44	327.5	14.6	301150	1	AP004602	Oceanobac	AP004602 Oceanobac
45	324.5	14.4	6854	1	BSU392330	Bacillus sp	BSU392330 Bacillus sp

ALIGNMENTS

RESULT 1

AX839476
LOCUS AX839476 1305 bp DNA linear PAT 15-DEC-2003
DEFINITION Sequence 2 from Patent EP1347044.
ACCESSION AX839476
VERSION AX839476.1 GI:39922766
KEYWORDS Bacillus sp. KSM-KP43
SOURCE Bacillus sp. KSM-KP43
ORGANISM Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE Okuda, M.K., Sato, T.K., Saito, K.K., Sumitomo, N.K., Izawa, Y.K.,
AUTHORS Saeki, K.K., Kobayashi, T.K. and Nomura, M.K.
TITLE Alkaline protease
JOURNAL Patent: EP 1347044-A 2 24-SEP-2003;
Kao Corporation (JP)
FEATURES
Location/Qualifiers
1..1305
/organism="Bacillus sp. KSM-KP43"
/mol_type="unassigned DNA"
/db_xref="taxon:109322"
<1..1305
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAE85146.1"
/db_xref="GI:39922767"
/translation="NDVARGIVKADYVQSSVGLYQGOIVAVADTGLDTRNDSSMHE
APRKITADYALGRINNDNGTHVAGSVLNGSTNGKMGAPCANLVFOSIMDSGG
GLGGLPSLIQTLFSQAKAIHTNSWGAAYNGAYTDSRVDDYVRKNDMTILFPA
GNEGNGTITLSARSAKAIIVGTANTENRPSFYADININHFSSRPGTKDRIK
DYVAPGTFLISARSLADPSFWANHSKYAYMGTSNATPIVAGNVAQLREHFVKR
GITPKPSLLKALIAAGADIGLYPGNGQWGRVTLDKSLNVAYNVSSLSLSTOKAT
YSFATAGKPLKISLVMSDAPASTTASVTLVNDLVLITAPNGTIVYVGNDFTSFYNDN
WDGRNVENVFINAQSGTYTIEVQAYNVPGVPQFSLAIVN"
ORIGIN
Alignment Scores:
Pred. No.: 1,71e-133 Length: 1305
Score: 2247.00 Matches: 434
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-985-689A-1 (1-434) x AX839476 (1-1305)
Qy 1 AsnAspValAlaAspGlyLeuValAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1 AATGATGTCGCGTGGAAATGTCGAAGCGGATGTGGCTCAGAGCAGCTACGGTGTAT 60
Qy 21 GlyGlnGlyGlnLeuValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GGACAAGGACAGATCGTAGCGGTGCGGATACAGGCTTGATACAGTCCGAATGACAGT 120
Qy 41 SerMetHisGlnAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCATGATGATGAAGCCCTCCGCGGGAATTTACTGCAATTATATGCAATGGGACGGAAT 180
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 181 AATGCCAATGATACGAATGTCATGTCATGTCATGTCGCTGGCTCCGATTTAGGAAACGCG 240
Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 241 TCCACTAATAAAGAAATGGCGCTCAGCGGAATCTAGTCTTCCAATCTATCATGTGATAGC 300
Qy 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 301 GGTGGGGGACTTGGAGGACTACCTTGAATCTGCAACCTTATTGACGCAAGCATACAGT 360
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyValAlaValAsnGlyAlaTyrThrThr 140
Db 361 GCTGGTGCCAGAAATTCATAAACTCTGCGGGAGCAGCAGTGAATGGGCTTACACAACA 420

Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 421 GATTCCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCCTTTTGGTGGC 480
Qy 161 GlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 481 GGAATGAAGGACCGGAACCGGACCATCAGTGCACAGGACAGCTAAAATGCAATA 540
Qy 181 ThrValGlyValAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 541 ACAGTCGGAGCTACGGAACCTCCGCCAAGCTTTGGTCTTATCGGACCAATATCAAC 600
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 601 CATGTGGCAGTCTCTTCAGTGCACGCAAGGATGGACGATCAACCGGATGTC 660
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
Db 661 ATGGCACCGGAACTTCTATCTATCAGCAAGATCTTCTCTTGACCGGATTCCTCTTC 720
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 721 TGGGCGAAACCATGACAGTAAATATGATGATGCGTGGTGGAACTCCATGCTACCGGATC 780
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 781 GTTGTGGAAACCTGGCACAGCTTCGTGAGCATTTTGTGAAAAACAGAGGCAATCACCCA 840
Qy 281 LysProSerLeuLeuLysAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 841 AAGCTTCTCTATTAAAGCGGCACTGATTCGCGTGGCAGCTGACATCGGCTTGGCTAC 900
Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 901 CCGAACGTTAACCAAGATGGGACGAGTGACATTTGGATAAATCCTTGACCTTGCTAT 960
Qy 321 ValAsnGlnSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 961 GTGAACGAGTCCAGTCTCTATCCACGACCAAAAGCGACGTACTCGTTTACTGTACT 1020
Qy 341 AlaGlyLysProLysLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 1021 GCGGGCAAGCCTTTGAAAATCTCCCTGGTATGTGTGATGCCCTCGGAGCACAATGCT 1080
Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 1081 TCCGTAAACGTTGTCAATGATCTGGACCTTGTCTATTACCGCTCCAAATGGCACAGTAT 1140
Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
Db 1141 GTAGGAATGACTTACTTCGCCATACAATGATACTGGGATGGCGCAATAACGTAGAA 1200
Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1201 AATGATTTAATATGACCAACAAAGCGGACGCTATACATTGAGGTACAGGCTTATAAC 1260
Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1261 GTACCGTGGACACACAGACCTTCTCGTTGGCAATTGTGAAT 1302

RESULT 2

AB051423
LOCUS AB051423 1923 bp DNA linear BCT 10-MAY-2002
DEFINITION Bacillus sp. KP43 PROF gene for protease, complete cds.
ACCESSION AB051423
VERSION AB051423.2 GI:20521154
KEYWORDS
SOURCE Bacillus sp. KSM-KP43
ORGANISM Bacillus sp. KSM-KP43
REFERENCE Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
AUTHORS Itoh, S. and Saeki, K.

TITLE new protease
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1923)
 AUTHORS Saeki, K.
 TITLE Direct Submission
 JOURNAL Submitted (21-NOV-2000) Katsuhisa Saeki, KAO CORPORATION;
 2606, AKABANE, ICHIKAIMACHI, HAGA, TOCHIGI 321-3486, Japan
 (E-mail: 387185@kao.co.jp, Tel: 81285687471 (ex. 7471),
 Fax: 81285687403)
 COMMENT On May 9, 2002 this sequence version replaced gi:14164344.
 FEATURES
 Location/Qualifiers
 1..1923
 /organism="Bacillus sp. KSM-KP43"
 /mol_type="genomic DNA"
 /strain="KP43"
 /db_xref="taxon:109322"
 1..1923
 /gene="PROF"
 1..1923
 /gene="PROF"
 /codon_start=1
 /transl_table=11
 /product="protease"
 /protein_id="BA55674.2"
 /db_xref="GI:20521155"
 /translation="MRKKKKVFLSVLSAAAILSTVALSNPSAGGARNFOLDPKGIQTT
 TDAGFSKQCGTGAALFVSENVKLPKGLQKLETPANNKLIHIFQNGPILIEETKQ
 QLEKTAGKIDLYTPDYAVIYVEYGDVKSATSTTEHVESVEPLYIYRIDPQLFTKAS
 ELVKAVADLPKQNKVQLRGIRQIAQFAISNDVLYITAKPEVKWMDVARGIVKADV
 AQSYGLYGGQIVAVADTGLDTRNDSSMHEAFRKITLALYALGHTNNNDTNGHGT
 HVAGSVLNGSTNGKNAQNALVFQSIMSGGLGGLPSNLQTLFSQATSAGARHIN
 SWGAENVGATIDSRNVDDIVRKNDMTILFAAGNEFGNGTISAPGTAKNAITVGATE
 NLRPSFGYADNINHAQVDSRPGTKDKRIKPDVMAPGTFILSARSLAPDSFWANH
 DSKYVNGTSMATPIVAGNVAQLREHFVKNRGTITPKSLKALIAAGADICLGYPN
 GNOGWGRTLDKSLNAVYNVESLSLTSOKATYSFTATAGKPLKILVWSDAPASPTA
 SVTLNLDLIVITAPNGTQVVGNDFTSPYNDNDGNNVNFVFNAPQSGTITIEVQA
 YNPVPGPQTFSLAVN"
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,76e-133 Length: 1923
 Score: 2247.00 Matches: 434
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 1 Gaps: 0
 US-09-985-689A-1 (1-434) x AB051423 (1-1923)
 Qy 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 619 AATGATGTTGGCGGTGAATTTGCAAGCGGATGTGGCTCAGAGCAGTACGGGTTGTAT 678
 Qy 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 679 GGACAAGGACAGATCGTAGCGGTGGCGATACAGGCGTTGATACAGTCGCAATGACAGT 738
 Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 739 TCGATGATGAAGCCTTCCCGCGGAAATTTACTGCATTATATCATTTGGACCGCGCAAT 798
 Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 799 AATGCCAATGATACGAATGTCATGTTACGATGCGATGTGGCTCGCTCCGATTATAGGAACGCG 858
 Qy 81 SerThrAsnIysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 859 TCCACTAATAAGGAATGGCGCTCAGCGCAATCTAGTCTTCCAACTCTATCATGGATAGC 918
 Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 919 GGTGGGGGACCTTGGAGGACCTACCTTCGAATCTGCNAACCTTATTACGCCNAGCATACAGT 978

Qy 121 AlaGlyAlaAspGlyIleHisThrAsnSerTyrGlyAlaAlaValAsnGlyValaTyrThr 140
 Db 979 GCTGTGCCAGAAATTCATCAAACTCTCTGGGAGCAGCAGTGAATGGGGCTTACACAA 1038
 Qy 141 AspSerArgAsnValAspAspTyrValArgIysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1039 GATTCCAGAAATGTGGATGATTAATGTGCGCAAAATGATATGATGATGATGATGATGAT 1098
 Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaIysAsnAlaIle 180
 Db 1099 GGGAAATGAAGACCGGACCGCGGAACCATCATGTCACCGAGGACAGCTAAAGTGAATA 1158
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 1159 ACAGTCGGAGCTACGGAAACCTCCGCCCAAGCTTGGTCTTATGGGACAATATCAAC 1218
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrIysAspGlyArgIleLysProAspVal 220
 Db 1219 CATGTGGCACAGTTCTCTTCACGTGGACCGCAAGGATGACGATCAAAACCGGATGTC 1278
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerLeuAlaProAspSerSerPhe 240
 Db 1279 ATGGCACCGGAAACGTTCTACTATCATCATGATGATGATGATGATGATGATGATGAT 1338
 Qy 241 TrpAlaAsnHisAspSerSerIleTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
 Db 1339 TGGCGGAAACCATGACAGTAAATATGATATGATGATGATGATGATGATGATGATGAT 1398
 Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIysAsnAspGlyIleThrPro 280
 Db 1399 GTTGTGGAAACGTTGGCACAGCTTCGTGAGCATTTTGTGAAAACAGAGGATCACACCA 1458
 Qy 281 LysProSerLeuLeuIysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 1459 AAGCCTTCTCTATTAAAAGCGCACTGATGTCGCGTGAGTGATGATGATGATGATGAT 1518
 Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspIysSerLeuAsnValAlaTyr 320
 Db 1519 CCGAAACGGTAACCAAGGATGGGACGAGTGATGATGATGATGATGATGATGATGATGAT 1578
 Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnIysAlaThrTyrSerPheThrAlaThr 340
 Db 1579 GTGAACGAGTCCAGTTCTCTATCCACGACCAAAAGCGAGTACTCGTTTACTGCTACT 1638
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 Db 1639 GCCGCAAGCCTTTGAAAATCTCCCTGGTATGGTCTGATGCGCTGGAGCACCACTGCT 1698
 Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1699 TCCGTAAACGCTTGTCAATGATCTGGACCTTCTCATTTACCGCTCCCAATGGCACAGTAT 1758
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 Db 1759 GTAGGAAATGACTTTTACTTCCCATACAATGATACTGGATGGGATGGCGCAATTAACGT 1818
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrThrIleGluValGlnAlaTyrAsn 420
 Db 1819 AATGATTTATTATATGACCAAAAGCGGACGATATACAAATGAGGTACAGGCTTATAAC 1878
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1879 GTACCGGTTGGACACACAGACCTTCTCGTTGGCAATTTGTGAAT 1920
 RESULT 3
 AR368117 LOCUS 1923 bp DNA linear PAT 12-SEP-2003
 DEFINITION Sequence 5 from patent US 6376227.
 ACCESSION AR368117
 VERSION AR368117.1 GI:34601778
 KEYWORDS Unknown.
 SOURCE Unknown.
 ORGANISM Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 1923)
 AUTHORS Takaiwa,M., Okuda,M., Saeki,K., Kubota,H., Hitomi,J., Kageyama,Y.,
 Shikata,S. and Nomura,M.

TITLE Alkaline protease

JOURNAL Patent: US 6376227-A 5 23-APR-2002;

FEATURES Location/Qualifiers

1..1923

source /organism="unknown"

/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
 Pred. No.: 2,76e-133 Length: 1923
 Score: 2247.00 Matches: 434
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-985-689A-1 (1-434) x AR368117 (1-1923)

Qy 1 AsnAspValAlaArgGlyLeuValLysAlaAspValAlaGlnSerSerThrGlyLeuTyr 20
 Db 619 AATGATGTTGGCGTGGAAATTTGCAAGCGGATGTGGCTCAGAGCAGCTACCGGTTGTAT 678
 Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrClyArgAspSer 40
 Db 679 GGACAGGACAGATCGTAGCGTTGCCGATACAGGCTTGATACAGGTGCAATGACAGT 738
 Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 739 TCGATGCATGAAGCCCTCCGCGGAAATTTACTGCATTATATGCTATGGACGGAGAT 798
 Qy 61 AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 799 AATGCCAATGATACGAATGTGTAGTCAGTGTGGCTGGCTCCGCTATTAGGAACGGC 858
 Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 859 TCCACTAATAAAGAAATGGCGCTCAGGCGAATCTAGTCTTCCAATCTATCATGATAGC 918
 Qy 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 919 GGTGGGGGACTTCGAGGACTACCTTCGAATCTCGAAACCTTATTCAGCCAAACATACAGT 978
 Qy 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAsnGlyAlaTyrThrThr 140
 Db 979 GCTGGTGCCAGATTTCATCAAACTCTGGGAGCAGCAGTGAATGGGCTTACACACA 1038
 Qy 141 AspSerArgAsnValAspSerTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1039 GATTCCAGAAATGTGGAGTACTATGTGCGCAAAATGATATGACGATCCTTTTCGCTGCC 1098
 Qy 161 GlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 1099 GGGAAATGAAGGACCGAAGCGGGAACATCATGTGACACAGGACAGCTTAAAAATGCAATA 1158
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 1159 ACAGTCGGAGTACGGAAACCTCCGCCCAAGCTTTGGGTCTTATCGGACAATATCAAC 1218
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 1219 CATGTGGCAGTCTCTTTCACGTGGACCGCAAGATGGAGCGGATCAACCCGGATGTC 1278
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
 Db 1279 ATGGACCGGGAACGTTTACTATCATCAGCAAGATCTTCTTTCGACCGGATTCCTCCITC 1338
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db 1339 TGGGCGAACCATGACAGTAAATATGATATCATGTCATGGGTGGAAACGTCCTACCCGATC 1398

Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 1399 GTTGCTGAAACGTGGCACAGCTTCGTGAGCATTTTGTGAAAAACAGAGGCATCACACCA 1458
 Qy 281 LysProSerLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 1459 AAGCCTTCTCTATTAAAGCGGCACACTGATTCCGGTGACGTGACATCGGCCCTTGGCTAC 1518
 Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 1519 CCGAACGTTAACCAAGATGGGACGAGTGACATTGGATAAATCCCTGAACGTTGGCTAT 1578
 Qy 321 ValAsnGlnSerSerSerSerSerThrSerGlnLysAlaAlaThrTyrSerPheThrAlaThr 340
 Db 1579 GTGAACGAGTCCAGTCTCTCTATCCACAGCCAAAAGCGACGTACTCTGTTACTGTACT 1538
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAla 360
 Db 1639 GCCGCAAGCCTTTGAAATCTCCCTGGTATGCTCTGATGCCCTCGGAGCACACTGCT 1698
 Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1699 TCCGTAAACGTTGCAATGATCTGGACCTTGTCTATTACCCCTCCAAATGSCACACAGTAT 1758
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTyrAspGlyArgAsnValGlu 400
 Db 1759 GTAGGAATGACCTTACTTCGCCATACATGATTAACCTGGATGCCCATACGTAGAA 1818
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1819 AATGATTATTATTATGACACACAAAGCGGACGTATACATTTAGGTACAGGCTTAAAC 1878
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1879 GTACCGTTGACACACAGACTTCTCGTTGGCAATTGTGAAT 1920

RESULT 4

AB084155 1923 bp DNA linear BCT 02-SEP-2003
 LOCUS Bacillus sp. KSM-9865 gene for protease, complete cds.
 DEFINITION
 AB084155
 ACCESSION
 AB084155.1 GI:34392386
 VERSION
 KEYWORDS
 SOURCE

Bacillus sp. KSM-9865

Bacillus sp. KSM-9865

Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

1

Okuda,M., Saeki,K. and Kobayashi,T.

Bacillus sp. KSM-9865 protease gene

Published Only in Database (2003)

2 (bases 1 to 1923)

Okuda,M., Saeki,K. and Kobayashi,T.

Direct Submission

Submitted (18-APR-2002) Mitsuyoshi Okuda, Kao corporation,

Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,

Tochigi 321-3497, Japan (E-mail:okuda.mitsuyoshi@kao.co.jp,

Tel:81-285-68-7543, Fax:81-285-68-7547)

Location/Qualifiers

1..1923

/organism="Bacillus sp. KSM-9865"

/mol_type="genomic DNA"

/strain="KSM-9865"

/db_xref="taxon:192495"

1..1923

/codon_start=1

/transl_table=11

/product="protease"

/protein_id="BAC82522.1"

/db_xref="GI:34392387"

/translation="MRKKKVFSLVSAAILSTVALSNPSAGSARNFOLDPKGIOTI

TDKGFQSKQTGAAPLVESENKLPKGLQKLETPANKLHIYQFNGPILERTKQ

OLEKTKAILDYPVAYIVEYEGDVKSATSTIEDVESVEPYLPYTRIDPQLFKGAS

FEATURES

source

1..1923

/organism="Bacillus sp. KSM-9865"

/mol_type="genomic DNA"

/strain="KSM-9865"

/db_xref="taxon:192495"

1..1923

/codon_start=1

/transl_table=11

/product="protease"

/protein_id="BAC82522.1"

/db_xref="GI:34392387"

/translation="MRKKKVFSLVSAAILSTVALSNPSAGSARNFOLDPKGIOTI

TDKGFQSKQTGAAPLVESENKLPKGLQKLETPANKLHIYQFNGPILERTKQ

OLEKTKAILDYPVAYIVEYEGDVKSATSTIEDVESVEPYLPYTRIDPQLFKGAS

CDS

1..1923

/organism="Bacillus sp. KSM-9865"

/mol_type="genomic DNA"

/strain="KSM-9865"

/db_xref="taxon:192495"

1..1923

/codon_start=1

/transl_table=11

/product="protease"

/protein_id="BAC82522.1"

/db_xref="GI:34392387"

/translation="MRKKKVFSLVSAAILSTVALSNPSAGSARNFOLDPKGIOTI

TDKGFQSKQTGAAPLVESENKLPKGLQKLETPANKLHIYQFNGPILERTKQ

OLEKTKAILDYPVAYIVEYEGDVKSATSTIEDVESVEPYLPYTRIDPQLFKGAS

ELVKAVALLDNTOKNKEVOLREGIOIAQFATNSDVLVITAKPEVKVMDVARGIVKADY
AQSYGLYGCCQIVAVADTGLDTRGNDSSMHEAFKGNKITALFALGRNNANDINGHT
HAGSVLNGSTNGKAPQANLVFQSIMDSGGGLGPNLQTLFQAYSAAGARLHTN
SWGAANGAYTTDSRNVDDYVRKNDMTILFAAGNEPGNGTISAPGTAKNAVITGATE
NLSPSGVADNINHVQFSSRGKTDGRIKPDVMAFGTIFLSASLSAPDSFWMANH
DSKYAYMGTSMTPTVAGNVAQSRHFKVNRGITPKPSLLKAALJAGAADIQLGYPN
GNQGWGRVTDKSLNVAYNVESLSGTOKATVSTATAGKPKLISLWSDAPASTTA
SVTLNDLDLIVITAPNGTQVGNDFSPYNNWNGRNVNFINAPQSGTITIEVQA
YNYVPVGPQFSLAIVN"

ORIGIN

Alignment Scores:

Pred. No.: 5,72e-133 Length: 1923
Score: 2242.00 Matches: 433
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.78% Indels: 0
DB: 1 Gaps: 0

US-09-985-689A-1 (1-434) x AB084155 (1-1923)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerTyrGlyLeuTyr 20
Db 619 AATGATGTTGACGTGGAATGTCAAAGCGGATGGCGCAGACGCTACGGGTTGTAT 678
Qy 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GGACAGGACAGATCGTAGCGGTGCGGATACAGGCGTTGATACAGGTGCGCAATGACAGT 738
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 739 TCGATGTCATGAAGCCITCCGGGGGAAATATACGTATTATGATTTGGGACCGGCGAAT 798
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 799 AATGCCAATGATACGATGTCATGTCACGATGGCTCGCTCCGTATTAGGAACGGC 858
Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 859 TCCACTAATAAGAAATGGCGCTCAGCGGAATCTAGCTTCCAACTCTATCATGGATAGC 918
Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 919 GGTGGGGGACTTGAGAGACTACCTTCGARTCGCAACCTTATTCAGCCAGCATACAGT 978
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAsnGlyAlaTyrThr 140
Db 979 GCTGTCGCCAGATTCATACAAACTCCTGGGGAGCAGCAGTGAATGGGCTTACACAA 1038
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1039 GATTCGAGAAATGEGATGACTATGTCGCAAAATGATATGACATCCCTTTTCGGTGCC 1098
Qy 161 GlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1099 GGGAAATGAAGACCGAAGCGGGAACCATCAGTGCACCGGCACAGCTAAATAATGCAATA 1158
Qy 181 ThrValGlyAlaThrGluAsnLeuValGlyProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1159 ACAGTCGGAGCTACGGAAACCTCCGCCAAGCTTCGGGTCTTATCGGGAACAATATCAAC 1218
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1219 CATGTGGCACAGTTCTCTTCACGTGGACCGCAAGAGATGACGGATCAACACCGGATGTC 1278
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 1279 ATGCAACGGGAACGTTTCATACACTATCAGCAAGATCTTCTCTGACCGGATTCCTCTTC 1338
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 1339 TGGCGGAACCATGACAGTAATATGCAATACATGGGTGGAACGTCCTACGCCATC 1398

Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGlyHisPheValLysAsnArgGlyIleThrPro 280
Db 1399 GTTGTGGAACCGTGGCACGCTTCGTGAGCATTTTGAAAAACAGAGGCATCACACA 1458
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 1459 AAGCCTTCTCTATTAAAAAGCGGCACTGATTCGCGGTGCAGCAGACATCGGCTTGGCTAC 1518
Qy 301 ProAsnGlyAsnGlnGlyTyrGlyValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 1519 CCGAACCGTAACCAAGGATGGGACGAGTGATTTGGATAAATCCTTAAACGTGGCTTAT 1578
Qy 321 ValAsnGluSerSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 1579 GTGAACGAGTCCAGTTCCTCTATCCACGACCCAAAGACGAGTACTCGTTTACTGCTACT 1638
Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
Db 1639 GCCGCAAGCCCTTTGAAGATCTCCCTGGTATGGTCTGATGCCCTGGCAGACCACTGCT 1698
Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 1699 TCCGTACGCTTGTCATGACCTGGACCTTCGTCTTACCTCCCAATGGCAGCAATAT 1758
Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
Db 1759 GTTGGAAATGACTTTTACTTCGCCATACAAATAATACTGGGATGCCGCAATAAGCTAGAA 1818
Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1819 AATGATTTATTTAATGCGCCAAAGCGGAGCATATCAATGAGGTACAGGCTTATTAAC 1878
Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1879 GTACCGGTGGACACACACCTTCCTGTTGGCAATTTGTGAAC 1920

RESULT 5

AR368118 1923 bp DNA linear PAT 12-SEP-2003
LOCUS Sequence 7 from patent US 6376227.
DEFINITION AR368118
ACCESSION AR368118
VERSION AR368118.1 GI:34601779
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1923)
AUTHORS Takaiwa, M., Okuda, M., Saeki, K., Kubota, H., Hitomi, J., Kageyama, Y.,
Shikata, S. and Nomura, M.
TITLE Alkaline protease
JOURNAL Patent: US 6376227-A 7 23-APR-2002;
FEATURES Location/Qualifiers
source 1..1923
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 5,72e-133 Length: 1923
Score: 2242.00 Matches: 433
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 1
Query Match: 99.78% Indels: 0
DB: 1 Gaps: 0

US-09-985-689A-1 (1-434) x AR368118 (1-1923)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 619 AATGATGTTGCGGTGGAATTTGTCAAAGCGGATGGCTCAGAGCAGCTACGGGTTGTAT 678
Qy 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40

Db	679	GGACAAGGACMGATCGTAGCGGTTCGCGATACAGGGCGTTGATACAGTCCGCAATGACAGT	738
Qy	41	SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyzAlaLeuGlyArgThrAsn	60
Db	739	TCGATGCATGAAGCCTTCGCGGAAAATTAATGCTATATGCTATATGCTATGCGACGACGAAT	798
Qy	61	AsnAlaAsnAspThrAsnGlyHisGlyThrHisValalaglySerValleuGlyAsnGly	80
Db	799	AATCCCAATGATACGAATGGTCATGGTACGCATGTGGCTGGCTCCGTATTAGGAACGGC	858
Qy	81	SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer	100
Db	859	TCACATAATAAAGGAATGGCGCTCAGCGAAATCTAGTCTTCCAAATCTATCATGGATAGC	918
Qy	101	GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeupheSerGlnAlaTyzSer	120
Db	919	GGTGGGGGACTTGGAGACTACCTTCGAAATCTGCAAACTTATTCAGCCAAAGCATACAGT	978
Qy	121	AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyzThrThr	140
Db	979	GCTGGTGCCAGAAATTCATAAACTCTCTGGGAGCAGCAGTGAATGGGCTTTACACAA	1038
Qy	141	AspSerArgAsnValAspAspTyzValArgLysAsnAspMetThrIleLeuPheAlaAla	160
Db	1039	GATTCCAGAAATGGGATGACTATGTGCGGAAAAATGATAGCATGCTTTTCGTGCTGCC	1098
Qy	161	GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle	180
Db	1099	GGGAATGAAGACCCGACGGCGGAACCATCAGTCCACGACGACAGCTAAAAATGCAATA	1158
Qy	181	ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyzAlaAspAsnIleAsn	200
Db	1159	ACATCGGAGCTACGGAAAACTCCGCCCAAGCTTTGGGTCTTATGGGACAATATCAAC	1218
Qy	201	HisValalaglnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal	220
Db	1219	CATGTGGCACAGTTCTCTTCACGTGGACCCGCAAAAGGATGACGAGTCAAAACGGAGTGC	1278
Qy	221	MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe	240
Db	1279	ATGCAACGGGAAAGCTTCATACTATCAGCAAGATCTTCTCTTGCACCGGATCTCTCCTTC	1338
Qy	241	TrpAlaIleHisAspSerLysTyzAlaTyzMetGlyGlyThrSerMetAlaThrProIle	260
Db	1339	TGGCGCAACCATGACAGTAATAATGCTATACATGGGTGGAAAGCTCCATGGGTACACCGATC	1398
Qy	261	ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro	280
Db	1399	GTTGCTGGAAACGTGGCACAGCTTCGTGAGGATTTTGTGAAAAACAGAGGCATCACACCA	1458
Qy	281	LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyz	300
Db	1459	AAGCCTTCTCTATTAAAAAGCGGCACCTGATTCGCGGTGCAGCTGACATCGGCTTGGGTAC	1518
Qy	301	ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyz	320
Db	1519	CCGAACGGTAACCAAGGATGGGACGAGTGACATGGGATAAATCCCTGAACGTGGCTTAT	1578
Qy	321	ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyzSerPheThrAlaThr	340
Db	1579	GTGAACGAGTCCAGTTCTCTATCCACGACCCAAAAAGCAGCTACTCGTTTACTGCTACT	1638
Qy	341	AlaGlyLysProLeuLysIleSerIleuValTrpSerAspAlaProAlaSerThrThrAla	360
Db	1639	GCCGGCAAGCCTTTGAAAATCTCCCTGGTATGGTCTGTATGCCCTGGCAGCACAACTGCT	1698
Qy	361	SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyz	380
Db	1699	TCCGTAAAGCTTGTCATGATCTGGACCTTGTCATTACCGGCTCCAAATGGCACACAGTAT	1758
Qy	381	ValGlyAsnAspPheThrSerProTyzAsnAspAsnTrpAspGlyArgAsnAsnValGlu	400
Db	1759	GTAGAAATGACTTTTACTTCGCCATACATGATAACTGGGATGGCGGCAATTAACGTAGAA	1818

Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyThrIleGluValGlnAlaTyAsn 420
Db 1819 AATGTATTTTAAATGACACCAAGCGGACGTATACAATTGAAGTACAGGCTTAAAC 1878

Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1879 GTACCGGTGGACACACAGAACTTCTGTTGGCAATTGTGAAT 1920

RESULT 6
AB046403
LOCUS
DEFINITION Bacillus sp. 9860 PROA gene for protease, complete cds.
ACCESSION AB046403
VERSION AB046403.2 GI:20521152
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE 2 (bases 1 to 1920)
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES

1920 bp DNA linear BCT 10-MAY-2002

Bacillus sp. 9860 PROA gene for protease, complete cds.

GI:20521152

Bacillus sp. 9860
Bacillus sp. 9860
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and Horikoshi, K.
Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic Bacillus spp.: enzymatic properties, sequences, and evolutionary relationships
Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)

20568675
11118284
Saeki, K.
Direct Submission
Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
Biological Science Laboratory, Tschikamachi Akabane 2605, Haga,
Tochigi 321-3497, Japan (E-mail:tschi185@kscatnet.kao.co.jp,
Tel:81-285-68-7400, Fax:81-285-68-7403)
On May 9, 2002 this sequence version replaced gi:12381938.
Location/Qualifiers
1. .1920
source

Alignment Scores:		
Pred. No.:	3,096-129	Length:
Score:	2183.00	Matches:
Percent Similarity:	99.31%	Conservative:
Best Local Similarity:	99.31%	Mismatches:
Query Match:	97.11%	Indels:
DB:	1	Gaps:
		1920
		418
		3
		0
		0

US-09-985-689A-1 (1-434) x AB046403 (1-1920)

QY 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 616 AATGATGTGGCCAGAGGATATGTCAAAGCGGATGTGGCACAGCAGCAGCTACGGTTTGTAT 675
QY 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGACAAGGCCAGATTCGCGAGTTCGCGATATCGGATTCGGATACAGGAGAAACACACAGT 735
QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATCATGACAGCCCTCCGCGTAAATAACAGCACTATATGCACTGGTGGTGGAGAT 795
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 796 AATGCGAATGATACGAACCGTATGTCACCTCCCATGTGGCAGGTTCGGATATGGAATATGC 855
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 856 GCAACGAAATAAGGATGACACCTCAGCGMATCTGGTTTTCATCCATCATGATAGC 915
QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 916 AGTGGTGGGCTTGGAGGCTTGCCTCCAAATCTGCAACCTTATTCAGCCAGCATTCAGT 975
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140
Db 976 CGAGGTGCCAGATTCATCAAACTCTCGGGGGCGAGCGGTGATGGGGCCCTACAGCA 1035
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1036 GATTCAGGAATGTGGATGACTATGTAAGGAAAAATGATATGACGATCTTTTCGGGT 1095
QY 161 GlyAsnGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1096 GGGAAATGAAGCGCGAAGCGGTATCATCAGTGCACCTGGTACGGCTAAACGCCATA 1155
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1156 ACAGTGGCGCAACCGAAACCTCGCTCAAGCTTCGGTTCTATGCAGATATATTAAC 1215
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1216 CACGTTGCACAGTTCTCTTCCCGTCCCGACAAAAGATGGCGAATCAAGCTCATGTC 1275
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
Db 1276 ATGGCCCGAGGACATACATTTTATCAGCAAGATCTTCTTGCACCCGATTCCTCCTC 1335
QY 241 TrpAlaAsnHisAspSerIleTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
Db 1336 TGGCGGAATCATGACAGCAATATGCTATATGGGTGGAACGCTCCATGGCAACACCGAT 1395
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIysAsnArgGlyIleThrPro 280
Db 1396 GTTGGCGGGAATTTGCACAGCTCCGTCAGCATTTTGTGNAATAACAGAGATCACTCCT 1455
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 1456 AAGCCTTCCCTATTGAAAGCAGCTTTGANTGACAGTGTCTGCTATGTTGGATTGGGTAT 1515
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 1516 CCGAACGAAACCAAGATGGGCGCAGTGACCTCGATATAATCGTTGACCTGTGCTAT 1575
QY 321 ValAsnGluSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 1576 GTGAACGAATCCAGTGGCCTTCACTAGCCAAAGAGACATATACCTTTACTGCAAG 1635
QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 1636 GCGGGCAAGCCATTTGAATCTCCCTGGTATGTGGATGCCCTGCCCTGCAAGCACTACTGCT 1695

QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 1696 TCTGTAAACCTCGTCAATGATTTGGATTTCGTCATTACAGCACCAACGGAACAGATAT 1755
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
Db 1756 GTCCGGGAATGACTTCTCGACCCANTTGCACATACTGGATGGCCGCAATACAGTAGAA 1815
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1816 AATGTAATTTAATTCGCCCAAGTGGAAACATATACCATTCAGGTGCAAGCATATAAT 1875
QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1876 GTCCGGTTGGACCAACAAACTTCTCGTTGGCAATTTGTGAAC 1917
RESULT 7
AR368116 AR368116 1920 bp DNA linear PAT 12-SEP-2003
LOCUS AR368116 Sequence 3 from patent US 6376227.
DEFINITION AR368116
ACCESSION AR368116
VERSION AR368116.1 GI:34601777
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1920)
AUTHORS Takaiwa,M., Okuda,M., Saeki,K., Kubota,H., Hitomi,J., Kageyama,Y.,
Shikata,S. and Nomura,M.
TITLE Alkaline protease
JOURNAL Patent: US 6376227-A 3 23-APR-2002;
FEATURES Location/Qualifiers
Source 1..1920
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 3,09e-129 Length: 1920
Score: 2183.00 Matches: 418
Percent Similarity: 99.31% Conservative: 13
Best Local Similarity: 96.31% Mismatches: 3
Query Match: 97.15% Indels: 0
DB: Gaps: 6
US-09-985-689A-1 (1-434) x AR368116 (1-1920)
QY 1 AsnAspValAlaAlaArgGlyIleValIysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 616 AATGATGTGGCCAGAGGATATGTCAAAGCGGATGTGGCACAGCAGCAGCTACGGTTTGTAT 675
QY 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGACAAGGCCAGATTTGTGGCAGTTCGCGATATCGGATTCGGATACAGGAGAAACACAGT 735
QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATCATGACAGCCCTCCGCGTAAATAACAGCACTATATGCACTGGTGGTGGAGAT 795
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 796 AATGCGAATGATACGAACCGTATGTCACCTCCCATGTGGCAGGTTCGGTATTAGGAATGCG 855
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 856 GCAACGAAATAAGGATGACACCTCAGCGMATCTGGTTTTCATCCATCATGATAGC 915
QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 916 AGTGGTGGGCTTGGAGGCTTGCCTTCCAACTGCAACCTTATTCAGCCAGCATTCAGT 975
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140

976 GCAGGTGCAGAAATTCATACAAATCTCTGGGGGAGCGGTGAATGGGGCTACAGACA 1035
 Db
 141 ApsSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Qy
 1036 GATTCAGAAATGTGGATGACATATAGGAAATGATAGCATTTCTTTTCGGGCT 1095
 Db
 161 GlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Qy
 1096 GGAATGAAGGCGAAGCGGGTACCATCATGTCACCTGGTACGGCTAAAGACGCATA 1155
 Db
 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Qy
 1156 ACAGTCGGCGCAACCGAAACCTGGTCCAGCTTCGTTCCTATGCGAGATAATTAAC 1215
 Db
 201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Qy
 1216 CACGTTCACAGCTTCTCTCCGTGGCGGACAAAGATGGCGAATCAACGCTGATGTC 1275
 Db
 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Qy
 1276 ATGGCGCCAGGACATACATTTATCAGCAAGATCTTCTTGCAACCGATTCCTCCTTC 1335
 Db
 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Qy
 1336 TGGCGAATCATGACAGCAATATGCTATATGGTGGACGTCCATGGCAACCGATT 1395
 Db
 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Qy
 1396 GTTGGGGGAATGTTGCACAGCTCGTGAGCATTTTGTGAAATAATAGAGGAATCACTCCT 1455
 Db
 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Qy
 1456 AAGCCTTCCTATTGAAGACCTTTGATGACAGTGTCTGATGTTGGATTGGGTTAT 1515
 Db
 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Qy
 1516 CGAAACGGAACCAAGGATGGGGCGAGTGACCTGGATAAATGCTTGAACGTTGCCTAT 1575
 Db
 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Qy
 1576 GTGAAGAAATCCAGTGCCCTATCAATAGCCAAAGGACATATACCTTTACTGCAACG 1635
 Db
 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrAla 360
 Qy
 1636 CGGGCAAGCCATTTGAAATCTCCCTGGTATGTCGATGCCCTGCAAGCACTACTGCT 1695
 Db
 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Qy
 1696 TCTGTAACCTCGTCAATGATTGGATTGGTATTGATACAGCACCACCAACGGAAGATAT 1755
 Db
 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTyrAspGlyArgAsnAsnValGlu 400
 Qy
 1756 GTCCGGAATGACTTCTCAGCACCAATTTGACATAACTGGGATGGCGCGCAATACGTAGA 1815
 Db
 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Qy
 1816 AATGATTATTATTATTCGCCCCAAAGTGGAAACATATACCATTTGAGGTGCAAGCATATAAT 1875
 Db
 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Qy
 1876 GTGCCGGTGGACCAAACTTCGTTGGCAATTTGTGAC 1917
 Db

RESULT 8

AB046406
 LOCUS
 DEFINITION Bacillus sp. NV1 PROE gene for protease, partial cds.
 ACCESSION AB046406
 VERSION AB046406.1 GI:12381944
 KEYWORDS
 SOURCE Bacillus sp. NV1
 ORGANISM Bacillus sp. NV1
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

REFERENCE

AUTHORS

1 (sites)

Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and

Horikoshi, K.

Novel oxidatively stable subtilisin-like serine proteases from

alkaliphilic bacillus spp.: enzymatic properties, sequences, and

evolutionary relationships

JOURNAL Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)

MEDLINE 20568675

PUBMED 11118284

REFERENCE 2 (bases 1 to 1302)

Saeki, K.

Direct Submission

Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation, and

Biological Science Laboratory; Ichikaimachi Akabane 2606, Haga,

Tochigi 321-3497, Japan (E-mail: 387185@kasanet.kao.co.jp,

Tel: 81-285-68-7400, Fax: 81-285-68-7403)

FEATURES

Location/Qualifiers

1..1302

/organism="Bacillus sp. NV1"

/mol_type="genomic DNA"

/strain="NV1"

/db_xref="taxon:133781"

1..1302

/gene="PROE"

<1..>1302

/gene="PROE"

/codon_start=1

/transl_table=11

/product="protease"

/protein_id="BAB21269.1"

/db_xref="GI:12381945"

translation="NDVARGIVKDAQSSVGLYGGQVAVADTGLDTRNDSSMHE

APRGKTAIYALGRNTNANDPNHGTHVAGSVLNGTSKNGMAPQANLVFQSVDSNG

GLGGPSNVSTLPSGAYSAGAPIHTNSGAPVNGAYTDSRVNDVVRKNDMAVLFAA

GNEPGGTFISAPGAKNAITVGATENLRPSGYSADNINHVAQSSRGPTKDGRLKP

DVAAPGTFILSRSLAPDSFWANHSKIATMGTSMTPIVAGNVAGLRHFIRNR

GTIPESLKLAAALAGATDILGIGPSGQNGRVTLDKSLNVAFVNETSSLSLTKAT

YFTQSGKPLKISLWSDAPASVTLSVNDLVLITAPNGTKYGVNDFTPAYDNN

WDGRNVENVFINAQSGTYTVEVQAYNVQGPQAFSLAIVN"

ORIGIN

Alignment Scores:

Pred. No.: 6,49e-127 Length: 1302
 Score: 2143.00 Matches: 406
 Percent Similarity: 97.93% Conservative: 19
 Best Local Similarity: 95.55% Mismatches: 9
 Query Match: 95.37% Indels: 0
 DB: 1 Gaps: 0

US-09-985-689A-1 (1-434) x AB046406 (1-1302)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20

Db 1 AACGATGTTGCCAGAGCGCATTTGTAAAGCCGATGTTGCCAGAGCGCTATGGTTATAT 60

Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrClyArgAsnAspSer 40

Db 61 GGGCAAGGGCAAGCGTTCGAGTAGCGGATACCGGACTGGATACAGGGCGTATGACAGC 120

Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60

Db 121 TCATGATCATGACGCTTCCGAGGAAGATTACCGGATATATGCCCTTGGGAAGACAAAC 180

Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80

Db 181 AACGCCAATGATCCAAACGACACGAGCGCATGTTGCCGGATCTGTTTATAGCAACCGT 240

Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100

Db 241 ACTTCGAATAAGGATGGCTCCACAGACTACTTAGTTTCCATCTCTGTTATGACAGC 300

Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120

301 AATGGCGGACTTGGCGGACTGCTTCCAAATGTAAGTACATATTATTCAGCGAGCATATAGT 360
121 AlaglyValaArgIleHisThrAsnSerTrpGlyValaAlaValaAsnGlyAlaTyrThrThr 140
361 GCCGGTGGCCAGATCCATACGAATCATGGGAGCGCCTGTAAACGGAGCGTACACTACT 420
141 AspSerArgAsnValaAspAspTyrValaArgLysAsnAspMetThrIleLeuPheAlaAla 160
421 GATTCAGAAACGTAGACGATTATGTTGCTGTAATAATGATATGCGGCTCTTTTTCAGCG 480
161 GlyAsnGluClyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
481 GGTAAACGAGCGCGAATGCGGCACAAATCAGTCTCCTGGTACCAGGAATGCTATC 540
181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
541 ACAGTAGGGCAACAGAAAACCTGGGCCCAAGCTTTGGATCTTATGCTGACAAACATCAAT 600
201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
601 CATGTAGCACAGTTCCTCCCGGACCTTACAAAGGATGGAGCTATCAACCGGACGTA 660
221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
661 ATGGCACCAGCAATTTATTTATCGCAAGATCTTCTTGGCTCGGACCTCTCATTC 720
241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
721 TGGCAAAACCATGACACAAATATGCTTATATGGTGGAAACATCCATGGCGACACCGATT 780
261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
781 GTAGCTGGTAACGTTGGACAGTTACGTGAACATTTATCAAAAACAGAGAAATCATCTCT 840
281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
841 AAACCATCCTTGTGAAACAGCTCTTATTCGGGAGCAACTCATATCGGTCTTGGCTAT 900
301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
901 CCGAGTGGAAACCCAGAGTGGGAAGAGTAACATTTGACAAAGTCACTTAATGAGCTTTC 960
321 ValAsnGluSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
961 GTAAATGAAACAAGCTCTTTATCTACTTACCAAAAGGCTACGTATTCTATTCTAGCACAA 1020
341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
1021 TCAGGCAAAACCTTTGAAGATTTTCAATGGTTTGTGCTGATCCCGGCAAGTACTTCCGCA 1080
361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
1081 TCGGTTACATTTGGTGAATGATCTGGATCTGGTGAATACAGTCCAAATGGAAAGATAT 1140
381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
1141 GTTGGAAACGACTTACTGCTCCCTATGATAATAACTGGGATGGAGCTAACATGTAGAG 1200
401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
1201 AACGTGTTTATCAATGCTCCGCAAGCGGAACGTATACAGTTGAGGTACAGGCTTACAT 1260
421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
1261 GTTCCCAAGGGCGCGAGGGTTCCTTTCCTTGGCTATTGTGAAC 1302

RESULT 9
LOCUS AR069954 3003 bp DNA linear PAT 18-FEB-2000
DEFINITION Sequence 41 from patent US 5891701.
ACCESSION AR069954
VERSION AR069954.1 GI:7220842
KEYWORDS

SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3003)
AUTHORS Sloma,A. and Christianson,L.
TITLE Nucleic acids encoding a polypeptide having protease activity
JOURNAL Patent: US 5891701-A 41 06-APR-1999;
FEATURES Location/Qualifiers
source 1..3003
/organism="unknown"
/mol_type="unassigned DNA"
Alignment Scores:
Score: 2,34e-125 Length: 3003
Pred. No.: 2125,50 Matches: 406
Percent Similarity: 97.93% Conservative: 19
Best Local Similarity: 93.55% Mismatches: 8
Query Match: 94.59% Indels: 1
DB: 6 Gaps: 1
US-09-985-689A-1 (1-434) x AR069954 (1-3003)
Qy 1 AsnAspValAlaAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1470 AATGACGTGGCGCTGGCATTTGTGAAGCAGACGCTGCACAAAATAACTTTGGCTTATAT 1529
Qy 21 GlyGlnGlyGlnIleValAlaAlaValaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1530 GGACAAGACAGATTGTAGCAGTTGCTGATCTCTGGCTTGATACAGGAAGAAATGCACGT 1589
Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 1590 TCGATGCGATGAAGCATTCGCGGTAAAGATTACCCACTATATGACCTGGCGCAACGAAT 1649
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 1650 AACGCCAATGATCCAAATGGACATGGAAACCCATGCTGGTGTAGGAAT--- 1706
Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 1707 GCTCAAAATAAAGGATGGCACCCAGCAAGCCAAATCTAGTCTTTCAATCTATTATGGATGT 1766
Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 1767 GGTGGAGGGCTGGGAGGACTACCTGCTAATCTACAAATATTTCAGTCAAGCATATAGT 1826
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValaAsnGlyAlaTyrThrThr 140
Db 1827 GCTGAGCGAGAAATTCATACGAATTCATGGGGGCTCCAGTAAACGGTCCCTATACGACA 1886
Qy 141 AspSerArgAsnValaAspAspTyrValaArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1887 GACTCTCGAATGTTGATGATTATGTGAGAAAATAATGATATGACGATCTTTTTCGGGCC 1946
Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1947 GGAAATGAGGGACACAGGTAGCGGTACAAATCAGTCCAGCAAGCAAGCAAGCAAGCAAGT 2006
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 2007 ACAGTTGGGGCAACCGAAAACCTTACGTCCAAAGCTTCGATCTTATGGGATATATTAAC 2066
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 2067 CATGTTGCTCAATCTCTTCACGAGTCTACTAGAGTGGACGTATTAAGCCGAGCGT 2126
Qy 221 MetAlaProGlyThrPheIleSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 2127 ATGGCACCAGGTACGATATATTTCTCTGCTAGATCATCAATAGTCTCAGATATCTTCATTC 2186
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260

2187	Db	TGGGCAAA	CCGATAGTAATAATAGCTCATCGGTGGTACTTCTATGGCTACTCCAAATT	2246
261	QY	ValAlaGlyAsn	ValAlaGlnLeuA ⁺ GgiuHisPheValIysAsnArgGlyIleThrPro	280
2247	Db	GTAGCAGGTAATCTTGCACAATTAAGGGAGCATTTTGTGAAAAATAGAGGGTAATCTCT	2306	
281	QY	LysProSerLeuLeuLysAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr	300	
2307	Db	ALGCGTTCCCTTTTAAAGCTGCTTTAATTGCAGGTGCTCGGATGTGGACTTGCGTTT	2366	
301	QY	ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr	320	
2367	Db	CCAAATGGTAACCAAGGATGGGAAGAGTAACGTTAGATAAATCCCTAAATGTCGCATTTT	2426	
321	QY	ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr	340	
2427	Db	GTGAATCAAAACGAGCCCTTTATCAAACAGTCAAAAGACACATATTCGTTTACGGTCA	2486	
341	QY	AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProIaSerThrAla	360	
2487	Db	GCTGGTAAACCCCTTAAAAATATCACTGTTGGTCAGATGCACACAGGTAGCAGACGGCA	2546	
361	QY	SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr	380	
2547	Db	TCACTAACCTTAGTGATGATTTAGACTTAGTAAATCACTGCACCAAAATGGNACTAAATAC	2606	
381	QY	ValGlyAsnAspPheThrSerProTyrAsnAspAsnTyrAspGlyArgAsnAsnValGlu	400	
2607	Db	GTCCGAATAGACCTTACAGCACCGCTATGATAACAATGGGATGGCAGAAACAACCTGGAA	2666	
401	QY	AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn	420	
2667	Db	AATGTGTTTATCAATGCTCTCAAGCGGACGTATACAGTCGAAGTGCAGGGTTTACAT	2726	
421	QY	ValProValGlyProGlnThrPheSerLeuAlaIleValAsn	434	
2727	Db	GTACCAAGTAGTCGCAAAACCTTTCTTTAGCGATTGTACAT	2768	

Percent Similarity:	97.93%	Conservative: 19
Best Local Similarity:	93.55%	Mismatches: 8
Query Match:	94.59%	Indels: 1
DB:	6	Gaps: 1
US-09-985-689A-1 (1-434) x BD062155 (1-3003)		
Qy	1	AsnAspValAlaArgGlyLeuVallysAlaAspValAlaGlnSerSerTyrglyLeuTyz 20
Db	1470	AAATGACGTGGCGGTGGCATTTGTGAAGCAGACGTGCGCAAAATAACTTTGGCTTTATAT 1529
Qy	21	GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db	1530	GGACAGAGCAGATTGTAGCAGTTGCTGATCTGGCTTGTATACAGGAAGAAATGACAGT 1589
Qy	41	SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyzAlaLeuGlyArgThrAsn 60
Db	1590	TCGATCGATGAAGCATTCGCGGTAAAGATTACCGCACTATATGCACTGGGCAGAACGAAT 1649
Qy	61	AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db	1650	AACCCAAATGATCCAAATGGACATGGACCAATGTTGCTGGATCTGTGTAGGAAT--- 1706
Qy	81	SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db	1707	GCTACAAATAAAGGATGCGACGCAAGCAATCTAGTCTTTCAATCTATATGGATAGT 1766
Qy	101	GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyzSer 120
Db	1767	GGTGGAGGGCTGGAGGAGTACTCTGCTAATCTACAAACAAATTTTCAGTCAGGCATATAGT 1826
Qy	121	AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyzThr 140
Db	1827	GCTGGAGCGAGATTTCATCAGAAATTCATGGGGGCTCCAGTAAACGGTGCTATACGACA 1886
Qy	141	AspSerArgAsnValAspAspTyzValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db	1887	GACTCTCGAAATGTCATGATTATGTGAGAAAAAATGATATGACGATCTTTTTCGGGCC 1946
Qy	161	GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db	1947	GGAATATAGGGACCAGGTAGCGGTACAACTAGTCGACACGAGAACACGCAAAAATGCGATT 2006
Qy	181	ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyzAlaAspAsnIleAsn 200
Db	2007	ACAGTTGGGCGCAACCGAAACCTACGTCCAAGCTTCGGATCTTATCGGATATAATAAC 2066
Qy	201	HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db	2067	CATGTTGCTCAATTCCTTCACAGGTCTCTACAGAGTGGAGCTATTAAAGCCGACGTC 2126
Qy	221	MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db	2127	ATGSCACCGGTACGTATATTCTCTCTGTAGATCATCATAGTCCAGATTCTCATTC 2186
Qy	241	TrpAlaAsnHisAspSerLysTyzAlaTyzMetGlyGlyThrSerMetAlaThrProIle 260
Db	2187	TGGCGCAACCATGATAGTAATAATGCTACATGGGTGGTACTTCTATGGCTACTCCCAAT 2246
Qy	261	ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db	2247	GTAGCAGGTAACTTTCACAAATTAAGGAGCATTTTGTGAAAAATAGAGGGTAACCTCCT 2306
Qy	281	LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyz 300
Db	2307	AAGCCTTCCCTTTTAAAGCTGCTTTAATGTCAGGTGCTGCGGATGTGTGACATTGGCTTT 2366
Qy	301	ProAsnGlyAsnGlnGlyTrpGlyArgGlyThrLeuAspLysSerLeuAsnValAlaTyz 320
Db	2367	CCAAATGGTTAACAGATGGGGAAGAGTAACGTTTAGATAAATCCCTTAATGTGCGATT 2426
Qy	321	ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyzSerPheThrAlaThr 340

Db 2427 GTGAATGAACGAGCCCTTTATCAACAAGTCACAAAGCAACATATTCGTTACGGCTCAA 2486
 Qy 341 AlAGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 Db 2487 GCTGGTAACCCCTTAAATAATACCTTGTGTGGTCAGATGACACGAGTAGCAGCGCA 2546
 Qy 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 2547 TCACAACTTTAGTGAAATGAATTTAGATCTAGTATCACTGCACCAATGAACATAATAC 2606
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGlu 400
 Db 2607 GTCGGAATGACTTTACAGCACCGCTATGATAACAATTTGGGATGGCAGAAAACAACGTGAA 2666
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 2667 AATGTGTTTATCAATCTCCTCAAGCGGAAGCTATACAGTCGAAGTGCAGGCTTACAT 2726
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 2727 GTACCAAGTAAGTCGCAACCTTTCTTTAGCGATTGTACAT 2768
 RESULT 11
 AB046405
 LOCUS AB046405 1299 bp DNA linear BCT 23-JAN-2001
 DEFINITION Bacillus sp. SD521 PROD gene for protease, partial cds.
 ACCESSION AB046405
 VERSION AB046405.1 GI:12381942
 KEYWORDS
 SOURCE Bacillus sp. SD521
 ORGANISM Bacillus sp. SD521
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 REFERENCE 1 (sites)
 AUTHORS Saeki,K.; Okuda,M., Hatada,Y., Kobayashi,T., Ito,S., Takami,H. and
 Horikoshi,K.
 TITLE Novel oxidatively stable subtilisin-like serine proteases from
 alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
 evolutionary relationships
 JOURNAL Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
 MEDLINE 2056875
 PUBMED 11118284
 REFERENCE 2 (bases 1 to 1299)
 AUTHORS Saeki,K.
 TITLE Direct Submission
 JOURNAL Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
 Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
 Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp,
 Tel:81-285-68-7400, Fax:81-285-68-7403)
 Location/Qualifiers
 1..1299
 /organism="Bacillus sp. SD521"
 /mol_type="genomic DNA"
 /strain="SD521"
 /db_xref="taxon:133780"
 1..1299
 /gene="PROD"
 /gene="PROD"
 /codon_start=1
 /product="protease"
 /protein_id="BAB21268.1"
 /db_xref="GI:12381943"
 /translation="NDVARGIVKADVQNNYGLYGGQVAVADTGLDTRNDSMHE
 AFKRIKIALVALGRTNANDPNGHTGTVAGSVLGNALNKGMAQOANLVFOSIMDSGG
 LGGPISNLNTLFSQAWNAGARLHTNSWGAIPNGAYTANSFQVDEYVRNNDMTVLPAAG
 NEGNSGTSIAPGTAIXVATGATENYRPFGLADPNPHIAQFSRGATRDRIKRPD
 VTAPGTFILSRSLADPDSFWANYSKIATMGTSWATFVAGNVAQUREHFIKNG
 ITPKPSIKKALITAGTVGLVPSGQGWGRTLDKSLNVAIYNATLALATQKATY
 SFOAQAGKPLISLVTDAPGSTASYTLVNDLDELVITAPNGQKIVGNDPSYPIDNW
 DGRNNVENVFINAPQSQTYTIEVQAVNVPSPGQRFSLAIVH"
 ORIGIN

Alignment Scores:
 Pred. No.: 9, 02e-118 Length: 1299
 Score: 1998.50 Matches: 381
 Percent Similarity: 94.24% Conservatives: 28
 Best Local Similarity: 87.79% Mismatches: 24
 Indels: 1
 Gaps: 1
 DB: 1
 US-09-985-689A-1 (1-434) x AB046405 (1-1299)
 Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 1 AATGATGTAGCAAGAGGAATAGTAAAGACGACGTTGCACAAAACAATACGGACTATAT 60
 Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 61 GGACAAGGTCAGTAGTAGTTCAGTAGCGGATACGGGTTTAGATACAGGTCGTAACGATAGT 120
 Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 121 TCTATGCGATGAGCAATTCCTGGGAAATTAACAGCTCTTTACGGGTTAGGAAGAACTAAC 180
 Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 181 AATGCAATGATCCGAATGGGCATGCTAGCGCATGTAGCAGGTTCTGTACTTGGTAAT--- 237
 Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 238 GCTTTAAATAAAGGAATGGCTCCGCAAGCTAACTTAGTCTTCCCAATCTATTATGGATAGC 297
 Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 298 ACGGAGGATTAGGTGGATTACCATCGAATTTGAATAGCTATTATTAGTCAGAGCTGGAAAT 357
 Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyValTyrThrThr 140
 Db 358 GCTGGGGCTAGAAATTCATACTAACTCTTGGGGTGCTCCAGTAAATAGGAGCGGTACACTGCT 417
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 418 AACTCGACAAGTGGATGAGTATGTTGCAATAATGATATGACGGTACTTTTTCGACGA 477
 Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 478 GGTAAATGAAGTCCTTAATTCAGAAACAATAGTGTCTCCAGCAGCAGCAAAAATGCCATT 537
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 538 ACGGTCCGCGCAACGGAATACTATCGCCGAGCTTCGGTTTCATTAGCAGATAACCCAAAT 597
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 598 CATATCGCACAAATTTTCATCAAGAGAGCTACGAGAGATGGACGAATTAACACGACGTA 657
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 658 ACAGCTCTCTGGAAACATTTATTATTCAGCACGCTTCTTCTTAGCCCCAGACTCTTCGTTT 717
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db 718 TGGGCGAATTAACAGTAAGTATCGGTACATGGCGGTACTCTATGCGGACACCTATA 777
 Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 778 GTTGGGGGAATGTCGCGCAATTCAGCGAGCATTTTATAAAAAATAGAGGAATTAACACT 837
 Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 838 AAACCTTCTTAAATAAAGCTGCACTATCGCTGGGGCTACTGATGTTGGTTAGGATAT 897
 Qy 301 ProAsnGlyAsnGlnGlyTyrTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 898 CCAAGTGTGTGACCAAGGCTGGGGCGTGTACTTCTAGTATAAATCGTTAAATGTAGCGTAT 957

Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrThrSerPheThrAlaThr 340
 Db 958 GTCAATGAGCACTGATAGCAGACAGACAAAGCAACGATTCGTTCACACAA 1017
 Qy 341 AlaGlyLysProLeuLysSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 Db 1018 GCGGGTAAACCTTTAAATCTCGTTAGTAGTACAGACAGTGCACCTGGAGTAGTCAACTGCA 1077
 Qy 361 SerValThrLeuValAsnAspLeuValLysThrAlaProAsnGlyThrGlnTyr 380
 Db 1078 TCTATACACTAGTTAATGATTTAGATCTAGTTTACTGCTCCGAAATGCAAAATAT 1137
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 Db 1138 GTAGGAATGATTTAGTTATCTTATGATATAAATACCTGGATGCTCGCAACATGTTGAG 1197
 Qy 401 AsnValPheLeuAsnAlaProGlnSerGlyThrTyrThrLysGluValGlnAlaTyrAsn 420
 Db 1198 AACGTATTTATAACCTCGCAATCTGGAACGTATACCAATTGAGGTTCAAGCGTATAT 1257
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaLysValAsn 434
 Db 1258 GTACCATCTGGCCACAGCGTTTCTCACTAGCTATCGTACAT 1299

RESULT 12
 AB046402
 LOCUS
 DEFINITION Bacillus sp. D6 PROA gene for protease, partial cds.
 ACCESSION AB046402
 VERSION AB046402.1 GI:12381936
 KEYWORDS
 SOURCE
 ORGANISM
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and
 Horikoshi, K.
 Novel oxidatively stable subtilisin-like serine proteases from
 alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
 evolutionary relationships
 Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
 11118284
 2 (bases 1 to 1299)
 Saeki, K.
 Direct Submission
 Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
 Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
 Tochigi 321-3497, Japan [E-mail:387185@kasanet.kao.co.jp].
 Tel:81-285-68-7400, Fax:81-285-68-7403

FEATURES
 Location/Qualifiers
 1..1299
 /organism="Bacillus sp. D6"
 /mol_type="genomic DNA"
 /strain="D6"
 /db_xref="taxon:127889"
 1..1299
 /gene="PROA"
 <1..1299
 /gene="PROA"
 /codon_start=1
 /transl_table=11
 /product="protease"
 /protein_id="BAB21265.1"
 /db_xref="GI:12381937"
 /translation="NDVARGIVKADVAQNNYGLYGQGVVAVADTGLDTGRNDSWHE
 AFKCKTALYALGRNNDPNHGHTVAGVGLGNALNKMGAPQANLVPQSLMDSGG
 LGGLPNLNTLFSQWNAGARLHTNSGAPVNGAYTANRQVDEYVRNNDMTVLFAAG
 NEGNPGSTISAPTKNAITVGATNRPFSIADNPNIHQFSSRGATDGRILKPD
 VTAPGTFILSAPSLAPDSFWANYSKYVMGTSMATPIVAGNVAQLRHFIRKRG
 ITPKSLIKALITAGATDVLGVPDQGWGRVTLDKSLNVAIVNEATLITGQKATY
 SFQTAQKPLKISLVWTDAPGSTASVTLVNDLIVITAPNGQKVGNDFSYFDNNW

DGRNVNVEFINAPQSGTYTIEVQAVNVPSGPORFSLAIVH"

ORIGIN

Alignment Scores:
 Pred. No.: 1.62e-117 Length: 1299
 Score: 1994.50 Matches: 380
 Percent Similarity: 94.01% Conservative: 28
 Best Local Similarity: 87.56% Mismatches: 25
 Query Match: 88.76% Indels: 1
 Gaps: 1

US-09-985-689A-1 (1-434) x AB046402 (1-1299)

Qy 1 AsnAspValAlaAlaArgGlyLysValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 1 AATCATGTAGCAAGAGCAATAGTAAAGACAGAGCTTGCACAAACAATACCGACTATAT 60
 Qy 21 GlyGlnGlyGlnLysValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 61 GGCACAGGTCAAGTAGTTGCGATAGCGATACCGGTTTAGATACAGGTCTGTAACGATAGT 120
 Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 121 TCTATGATGAGCATTCCTGGGAAATACAGCTTTTACGGTTAGGAGACTAAC 180
 Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 181 AATGCAATGATCGAATGGCATGGTACGATAGTGTCTGTCTGTCTGTGTAAT--- 237
 Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 238 GCTTTAATTAAGAAATGGCTCCGCAAGCTAATTAGTCTTCCAACTATTATGGATAGC 297
 Qy 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 298 AGCGGAGGATTAGGAGGATTAACATCGAATTTAAATACGTTATTAGTCAAGCTTGAAT 357
 Qy 121 AlaGlyAlaArgLysHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140
 Db 358 GCTGGCCTGAGAAATCATCACTACTCTTGGGGGCCCAAGTAATATGGAGGCTACCTGCT 417
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 418 AACTCGAGACAAGTGGATGAGTATGTTCGAAACAATGATATGACGGTACITTTTTCAGCT 477
 Qy 161 GlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaLys 180
 Db 478 GGAAATGAAGACCTAACTCTGGAAACAATTAGCGCTCCAGGACAGCGAAAATGCCATT 537
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 538 ACGGTCGGCGCAACGGAAACTACCGACCAAGTTTGGTTCAATTGCGAGATAACCCCTAAT 597
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 598 CATATCGACAATTTTCATCGAGAGGAGCTACGAGAGATGCGACGAATTAACACGACGTA 657
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 658 ACAGCTCTCGAACATTTATATATCAGCAGCGCTCTCTTTAGCACCAGACTCTTCGTTT 717
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db 718 TGGCGGAATTTATAACAGTAGTATCGCTATATGGCGGTACCTCTCTATGGCGACACCTATA 777
 Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 778 GTTCGGGGAATGTCGCGCAATTACCGCAGCAATTTTATAAAATAGAGGAATATACACT 837
 Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspLysGlyLeuGlyTyr 300
 Db 838 AAACCTTCTTATAATAAAGCTGCACCTATCGCTGGGCTACTGATGTGTTGTTAGGATAT 897

QY 301 ProAsnGlyAenGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 898 CCAAGTGGTGACCAAGGCTGGGGCGGTACTTTAGATAAATCGTTAAATGTAGCGTAT 957
 QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 958 GTCAATGAAGCACTGCATTAACACAGGACAAAGCAACGATTCGTTCCAAACGCCAA 1017
 QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
 Db 1018 GCGGGTAAACCATTAATAATCTGTTAGTAGTGACAGATGCACCTGGAAGTACACAGCA 1077
 QY 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1078 TCTTATACACAGTAAATGATTTAGATCTAGTATTACTGCTCGAATGGACAAAATAT 1137
 QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 Db 1138 GTAGTAATGATTTAGTTATCTTATGATAATTAATGGGATGGTCCACACAAATGTTGAG 1197
 QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1198 AACGTATTATTAACGCTCCGAATCTCGAAGCTATACAAATTCAGGTTCAAGCGTATAAC 1257
 QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1258 GTTCACTGACACACAGCGTTCTCACTAGCTATCGTACAT 1299

RESULT 13
 AB046404
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 JOURNAL
 TITLE
 JOURNAL
 TITLE
 JOURNAL
 FEATURES
 source
 1. .1299
 /organism="Bacillus sp. Y"
 /mol_type="genomic DNA"
 /strain="y"
 /db_xref="taxon:133779"
 1. .1299
 /gene="PROC"
 <1. .>1299
 /gene="PROC"
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="protease"
 /protein_id="BAB21267.1"
 /db_xref="GI:12381941"
 /translation="NDVARGIVKADVAQNNYGLYGGQVVAVDLTGDRNDSMHAE
 AFRGKITALVALGRTNNAQDPNGHGTGTVAGSVLGNALNKGMAFCALIVFOSIMDSGG
 LGGLPSNLTLFSAQWAGARIHTNSWGAQPVNGAYTANSRQVDEYVRNNDMTVLFAAG

ORIGIN

Alignment Scores:

Pred. No.: 4,48e-117 Length: 1299
 Score: 1987.50 Matches: 379
 Percent Similarity: 94.01% Conservative: 29
 Best Local Similarity: 87.33% Mismatches: 1
 Query Match: 88.43% Indels: 1
 DB: 1 Gaps: 1

US-09-985-689A-1 (1-434) x AB046404 (1-1299)

QY 1 AsnAspValAlaAspGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 1 AATGATGTAGCAAGAGGATAGTAAAGCTGATGTTCACAAACAATACGGATTATAT 60
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 61 GGACAAGGTCAGTAGTGTGAGTACGACAGGCTTAGATACAGGTCGTAACGATAGT 120
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 121 TCTATGCAATGAAGCATTCGCGGGAATTCACAGCTCTTTACGCGTTAGGAAGAATAAT 180
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 181 AATGCGAGTGATCCGAATGGCATGCGACACATGTAGCAGGTTCTGTGTTGTAAT 237
 QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 238 GCTTTAAATAAAGGAATGGCTCCGCAAGCTAACTTAGTCTTCCATCTATTATGATGATAGC 297
 QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 298 ACGCGAGGATTAGTGGTGTACCATCGAACTTAATACGTTATTAGTCAAGCTTGGAA 357
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140
 Db 358 GCTGGAGCAAGAATTCATACATACTCTTGGGAGCCCGCAAGTAAATGGAGCGTACACTGCT 417
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 418 AACTCGAGCAAGTGGATGATATGTTTGAATAATATATGACGCTACTTTTTCACGCT 477
 QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 478 GGTAAATGAAGTCTTAATTCAGGAACAATTAGTGTCCAGGTACAGCGAATAATGCTATT 537
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 538 ACGGTGCGCGCAACCGAAACCTATCGCCCAAGCTTCGTTTCGATAGCAGATAACCCAAT 597
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 598 CATATTGCACAATTTTCATCGAGAGGAGCTACGAGGATGACGAATTAAGCCTGACGTA 657
 QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 658 ACAGCTCTCGGAACATTTATTTTATCAGCAGCTTCTTCCTTAGCTCCAGACTCTTCGTT 717
 QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db 718 TGGCGCAATTATAACAGTAATACCGGTATATGCGCGTACTCTCATGCGCACACTATT 777
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 778 GTTGCAAGGAATGTGCGCAATTAGCTGAGCATTTTATAAAAAATAGAGGTATTACTCTCT 837

QY 281 LysProSerLeuLeuLysAlaAlaLeuLleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 838 AAGCCTCTCTTTAATAAAGCTGACCTATCGTGGCTGCTACTGATGTTGGTTAGGATAT 897
 QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 898 CCTAATGGTGACCAAGCTGGGGCGTGTACTCTAAATAAATCGTTAAATGTCGGTAT 957
 QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 958 GTCATGAAAGCACTCATAGCCACAGGACAAAGCAACGTAATTCGTTCCAGCACAA 1017
 QY 341 AlaGlyLysProLeuLysLysLeuValTrpSerAspAlaProAlaSerThrThrAla 360
 Db 1018 GCGGTAACCTTTAAATCTCGTAGTAGGACAGATGCTCTCGGAAGTACACTGCA 1077
 QY 361 SerValThrLeuValAsnAspLeuAspLeuValLleThrAlaProAsnGlyThrGlnTyr 380
 Db 1078 TCTTATACACTAGTTAATGATTTAGATCTAGTTATTACTGCTCCGAATGGACAAATAT 1137
 QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 Db 1138 GTAGGAATGATTTTGTAGTTATCTTTATGATATACTGGATGGTGGCAACATGTTGAG 1197
 QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1198 AACGTATTTATAAACGCTCCGAATCTGGAACGTATATAATTAATGAGGTTCAAGCGTATAAT 1257
 QY 421 ValProValGlyProGlnThrPheSerLeuAlaLleValAsn 434
 Db 1258 GTACCATCTGGCCACAGCGTTTCTCAGTAGTCATCGTACAT 1299

RESULT 14

EO3808
 LOCUS 2218 bp DNA linear PAT 29-SEP-1997
 DEFINITION DNA encoding alkaline protease Ya.
 ACCESSION EO3808
 VERSION EO3808.1 GI:2172022
 KEYWORDS JP 1992197182-A/1.
 SOURCE Bacillus sp.
 ORGANISM Bacillus sp.
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 Tobe S., Odera M. and Asai Y.
 TITLE DNA CODING ALKALINE PROTEASE YA ENZYME AND PRODUCTION OF ALKALINE PROTEASE YA USING THE DNA
 JOURNAL Patent: JP 1992197182-A 1 16-JUL-1992;
 LION CORP

COMMENT OS Bacillus sp.
 PN JP 1992197182-A/1
 PD 16-JUL-1992
 PF 28-NOV-1990 JP 1990327110
 PI TOBE SEIICHI, ODERA MOTOYASU, ASAI YOSHIO
 PC C12N15/57,C11D3/386,C12N9/54,(C12N15/57,C12R1.07),(C12N9/54,C12R1.07);
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC *source: strain=y strain;
 CC *source: clone=pUB8A;
 FH Key Location/Qualifiers
 PH CDS 218..2125
 FT /product='precursor of alkaline protease Ya'
 FT sig_peptide 218..823
 FT mat_peptide 824..2122
 FT /product='Precursor of alkaline protease Ya'.
 FT Location/Qualifiers
 1..2218
 /organism="Bacillus sp."
 /mol_type="genomic DNA"
 /db_xref="taxon:1409"

FEATURES

source

ORIGIN

Alignment Scores:
 Pred. No.: 1.01e-116 Length: 2218
 Score: 1986.50 Matches: 379
 Percent Similarity: 94.01% Conservative: 29
 Best Local Similarity: 87.33% Mismatches: 25
 Query Match: 88.41% Indels: 1
 Gaps: 6

US-09-985-689A-1 (1-434) x EO3808 (1-2218)

QY 1 AsnAspValAlaAaCGGlyVileValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 824 AATGATGTAGCAGAGGGATAGTAAAGCTGATGTTGCACAAACAATACGATTATAT 883
 QY 21 GlyGlnGlyGlnLleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 884 GGACAAGGTCAACTAGTTGCGTAGCGGACACAGGCTTAGATACAGGTCGTAAACGATAGT 943
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 944 TCTATGATGATGAGCATTTCCGCGGGAATAATCAGAGCTCTTTACCGGTTAGGAAGACTAAT 1003
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 1004 AATCGAGTGCATCCGAATGGGCATGGCACACATGTAGCAGGTTCTGTACTTGGTAAT-- 1060
 QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 1061 GCITTAATAAAGAAATGGCTCCGCAAGCTAACTTAGTCTTCCACTATTATGGATAGC 1120
 QY 101 GlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 1121 AGCGGAGGATTAGTGCTTTACCATCGAAGCTTAAATACGTTATTTAGTCAAGCTGGAAT 1180
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
 Db 1181 GCTGAGCAAGAATTCATCTAATCTCTTGGGGAGCCCGAGTAATAGAGGCTACACTGCT 1240
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1241 AACTCGAGACAAAGTGGATGAGTATGTTGGAATAATGATGATGATGATGATGATGATGAT 1300
 QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 1301 GGTAAATGAGGTCCTTAATTCAGGAACAATTAGTCTCCAGGTACAGCAAAAATGCTATT 1360
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 1361 ACGTTCGGCGCACAGGAAACTATCGCCCAAGCTTCGGTTCGATAGCAGATACCCCAAT 1420
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 1421 CATATTGCACAAATTTTCATCGAGAGGAGCTACGAGGAGTACGAGGAATTAAGCCTGACGTA 1480
 QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 1481 ACAGTCTCTGGAACATTTATTTATCAGCAGCTCTCTCTTAGCTCCAGACTCTTCGTTT 1540
 QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
 Db 1541 TGGCGGAATTAACACAGTAATAACGCGTATATGGGGGTACCTCCATGGCGACACCTATT 1600
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 1601 GTTCAGGGAATGCGCGCAATTTACGTGAGCATTTTATAAAAATAGAGGTATTACTCCT 1660
 QY 281 LysProSerLeuLeuLysAlaAlaLeuLleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 1661 AAGCCTCTCTTTAATAAAGCTGCACCTTATCGCTGGTGTACTGATGTTGTTTAGGATAT 1720
 QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320


```
/db_xref="GI:9664581"
/translation="MSGNIFSKRHPAQIALIALISLLFVGTMQFIRSSHAAGGSDIDL
TVIIRDPENQAVVGLGILLFVYALITEVVRHTLAATLGSLIAVALIALNHSVEEAL
SUKAVTMDMETIGLLGMVWVGIISHTGFVFAVQYKSGGEITWILVLCVS
TAVSLAFNLTMLLTPTTIKQVLDLKPILISEVFSNTGGAATWIGDPNI
MIGASPDPAISGNKPNIANSQDVFENDEITEMAPGILMIVFSLMIRWLYKDE
FGRRDRDITELEQVAKIDIKGLKVGAILVLVILGPFHPVHILAVSWALSGALV
MLLATNHELEPLEHVEVETLLFFAGLPVLVHSLQYMGVIDYIGEYVEKAIKRPDD
YRLAAAILILMVSATASAPINIPVTATMIPIVLSLAPELNPLIMLALGACL
GGNGTLIGASNVYTAGMSEBAGYFISNEFRAGFPVWLLSTFIVSYFMYLIVYVGG
EGALTKWIKLVGISLGLIIAQYSGRAGKSPAEALVDDDFEIDFIVSTIKSVVTKTS
KGSSEEE"
complement(12019..12282)
/notes="37F11#9; contains KH domain"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="RNA-binding protein"
/protein_id="AAP97186.1"
/db_xref="GI:9664583"
/translation="MSVPNHIIRQAKDTSLPITMIGKSGLTDSVVVELEGQLSSRSL
VKVKNRGLYPRDELXTVMNIAEKTNSVVPARGNVAVFWEN"
complement(12279..12614)
/notes="37F11#10; contains Zn-finger; conserved protein"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="unknown"
/protein_id="AAP97185.1"
/db_xref="GI:9664582"
/translation="MGRSRRLNRDKKHAIWSQTCLLSLKNPQSOFTPEQSRFALH
LVKYSRRHRIPLSEVREIMCRITCNLLRYGENAIFRNGHGIOTCHSCTSVRKIPY
RQHSQKKA"
complement(12624..13274)
/notes="37F11#11"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="adenylate kinase"
/protein_id="AAP97187.1"
/db_xref="GI:9664584"
/translation="MDEGERQGMNLIVLFGPPGAGKGTQACRMMDATGLPOVSTGDM
EALKEIVESVAVLADVPDIVERICRGYSGCTGTVVHDFNFIQNGICGCSFTE
KRADDVEITVLEAREYHQTOTSPADFVRAAGIFHHVDDGDDIDIVITDILLVLG"
complement(13284..16691)
/notes="37F11#12"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="membrane-associated subtilysin-type serine
protease"
/protein_id="AAP97188.1"
/db_xref="GI:9664585"
/translation="MYMQUSGTQATAVAGGSASIAREYLREVAGINKPSASLIKATL
INGAELGDPIDIPNANGWQDLENSLNPSSGSLVDVQDDRELQAGSLIYSFD
LDGSKGIDITLWASDAEASANAQSERLNLNLDLILAPDGSVYLGNDFSGSITTC
GADNLNTERIRIPAGATTONGWVTVHRGGSSQRYSVIIRADATLIPKADLTTF
GNSILPSSSTPLIGDLVSLAWNQGLASQPVRIQEDITSGSILYDSNRSLGEG
SLDSVFFQFTPTTGVLKRLSLDTNNQVPELNDALINGIDNIDLEITAGQURVI
FQNCDSITQSDRSASVMDVRNETGLSLPFIHAHEGGERPDLSTVSSVQEPD
PYPTLLSPEDSWKSNQTVFTISGQGEENDIYTLNLDNTSASFQDQKRYAR
AGSFIVDLYARYQNTQVSHQVLYEIQGVDSVDFVPSGVSAGKERTGFSIGV
RNTGSAQYTMSCSTSSSQIMLGNSNSLSLEFPLNLOLSDMVNVVFPISANGE
PLAGSDTQVCTVPTSLNRVETDVAVSELKAFRSDLPDGPVGPALALPVF
ADTELAVFNHTIONKNVPLDFAVTLERGNPMAAEIQDFEQIISSTLSYTLGQGS
GDVERLLVPETAREGNSNTYLRVESPQMTLNTSLVGVGNLNVLSNVGTLIA
APVNDITFTEIREGNSNDLSEWSTSLPADGWSIGYSNPPSTVPSVLSQASVOLIA
KAPNTASQGFDMQDQVFNQTNNGRFTNABLLRNVLTAVATFAGISVDETIAPLFI
PRGDSQKQITITNEGNIPLSQLTVEVDGEGNLTDSSTISPESTDLAIGESIE
VVKVSTDESDIDGNMLVILITIDGVVIEFMADTVSSQSSGGIGFILPILSYVP
LVLLALLGYGRRLKQSSKDDGDTVELVADHTDADHLGTREQALDISHSVNDI
ASGEVSQDEIAAALASQDLPMPVTKKQVPTGTPSGLPVPGLPAGLPAGI
```

```
CDS
PPAKSYPLVPVQVTAGPPLPPGGLPDGWTMEQWNYGEQYLQRMGLN"
complement(16790..18130)
/notes="37F11#13"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="membrane-associated subtilysin-type serine
protease"
/protein_id="AAP97189.1"
/db_xref="GI:9664586"
Alignment Scores:
Pred. No.: 2,89e-29 Length: 60006
Score: 632.50 Matches: 166
Percent Similarity: 54.20% Conservative: 79
Best Local Similarity: 36.73% Mismatches: 148
Query Match: 28.15% Indels: 60
DB: 1 Gaps: 17
US-09-985-689A-1 (1-434) x AF268611 (1-60006)
Qy 13 AlaGlnSerSerTyrglyLeuTyrglyGlnGlyValAlaValAlaAaspThrGly 32
Db 17383 GCTCCTCTCTTCTTGAATCTTGATGGTCTCGGAACAGATTGCGATGGAATCTGGC 17324
Qy 33 LeuAaspThrGlyArgAaspSerSerMetHisGluAlaPheArgGlyIleThrAla 52
Db 17323 GTCGATATGGAT-----CATCCTGACCTCATCGACGGGTTCAGCA 17282
Qy 53 LeuTyrala---LeuGlyArgThrAsnAsnAlaAaspThrAsn---GlyHisGlyThr 70
Db 17281 ATTAACTACTGATTCGGTGTAGATAGTTCTCCAAATTGATAGCAACAGTGTGTCAGGACG 17222
Qy 71 HisValAlaGlySerValLeuGlyAsnGlySerThrAsn-----LysGlyMetAla 87
Db 17221 CACGTGCTATGACTGTGCTTGGCGATGTCACGGTGATGACGACGACGGGTGTGCT 17162
Qy 88 ProGlnAlaAsnLeuValPheGlnSerIleMetAspSerClyGlyLeuGlyLeu 107
Db 17161 CCAGAACCAATCTTGTGTCATGCTTTCAGCAGCACGACCAACCGGTGTCTTTGGAAGG 17102
Qy 108 ProSerAsnLeuGlnThrLeuPheSerGlnAlaty-SerAlaGlyAlaArgIleHisThr 127
Db 17101 CAAGTTTCATCTATGATCTCTCTGATGCCGAATTAACACAGCAAGTGTCCGAGTC 17042
Qy 128 AsnSerTrpGlyAlaAlaValAsn---GlyAlaTyThrThrAspSerArgAsnValasp 146
Db 17041 AATGCATGGGTCTTAAATGGAATTAACGAGCCTACACTGCTGATCA-AGATCGACAGAT 16983
Qy 147 AspTyrrValArgLysAsnAspMetThrIle---LeuPheAlaAlaGlyAsnGluGlyPro 165
Db 16982 CAATATGCTCCACATAAATAATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 16923
Qy 166 AsnGlyGlyThr---IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 184
Db 16922 ACAGTTTCTCAAAGATTACTGCTCCAGGAACACCAAGAAATGTTCTTCCATTTGGCTCA 16863
Qy 185 ThrGluAsnLeuArgProSerPheGlySerTyrrAlaAspAsnIleAsnHisValAlaGln 204
Db 16862 AGCCTTAAT-----GGTAGC-----GTCAGCTCC 16839
Qy 205 PheSerSerArgGlyProThrLysAspGlyArgIleLysProAspValMetAlaProGly 224
Db 16838 TATTTCATCGGAAGGTCCAATCTCGATGGAAGAATTAACCTCATTTAGTTCGCTCCCGGT 16779
Qy 225 ThrPheIleLeuSerAlaArgSerSerLeuAlaProasp----- 237
Db 16778 GACGGTATCTGTTCTCCAGCAGCAGGAGCAAGCAACGTTTGGTTCAGTCTCGCGGT 16719
Qy 238 SerSerPheTrpAlaAsnHisAspSerLysTyrrAlatyMetGlyGlyThrSerMetAla 257
Db 16718 ATTGGTACTCACTCAAAATGTCGTTCAATGATATATGATCAACTCAGCGGAACCTTCTCAAGCA 16659
```



```
Qy 258 ThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGly 277
Db 16558 ACACGATAGCAGGTGGTTT CAGCATCCCTCGACGTGAATATCTGCTGAAGTTGCTGGA 16599
Qy 278 IleThr---ProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIle 296
Db 16598 ATCAACAACCATCTGCAAGCCTCATTAAGCAACTCTCATCAACGGGGCAGAAGATTTA 16539
Qy 297 GlyLeu---GlyTyrProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSer 315
Db 16538 GGCACACGGACATTCCTTAAGCAACAGGGATGGGGGCAAAATTGATCTCGAAATATAGC 16479
Qy 316 LeuAsnValAla-----TyrValAsnGluSerSerSerLeu 327
Db 16478 CTGAATCCTTCGTCTGCTGGCGTTTCTCTGATGTCCTTCAGGATGATGAGCGTGAAC TT 16419
Qy 328 SerThrSerGlnLysAlaThrTyrSerPheThrAlaThrAlaGlyLysProLeuLysIle 347
Db 16418 CAGCAGGATTCTCATTTGATTTATCTCTTTGACCTTGATGGATCAAGGGTATTGACATT 16359
Qy 348 SerLeuValTyrSerAspAlaProAlaSerThrThrAlaSerValThr-----Leu 364
Db 16358 ACCTTGGCATGGACGGATGCAGAGCCAGTGCATAATGCTGCACAACTCTGAATCAGCTCTA 16299
Qy 365 ValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAsp 384
Db 16298 TTGAACAATTAGATTGATTTGATTTCTAAATTGCACCGGATGGGTCTTCATACCTTGGTAAATGAC 16239
Qy 385 PheThrSerProTyrAsn-----AspAsnTyrAspGlyArgAsnAsnValGluAsn 401
Db 16238 TTCTCTTCTGGGATTCAACACACAGCGGTTCCGCTGATTAATCTGAACATATCGAGCGC 16179
Qy 402 ValPheIleAsnAla-----ProGlnSerGlyThrTyrThrIleGluVal----- 416
Db 16178 ATTCGCATCCCTGCAGGAGCGACGACACAGAACGGTGAATTGGATGGTTACAGTTGAACAT 16119
Qy 417 -----GlnAlaTyrAsnValProVal 423
Db 16118 CGAGGAGGAGTTTCTCAGCGCTACGATCGTTATT 16083
```

Search completed: March 16, 2004, 00:36:45
Job time : 3658 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 15, 2004, 22:28:03 ; Search time 396 Seconds
(without alignments)
4655.853 Million cell updates/sec

Title: US-09-985-689a-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAVNPVGPOTFSLAIV 434

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-C=/cgn2.1/USPTO_epool/US09985689/runat_10032004_112902_19707/app_query.fasta_1.583
-DB=N Geneseq 29Jan04 -OFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRF=US09985689 @CGN.1 1 352 @runat_10032004_112902_19707 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=10 -YGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2247	100.0	1923	2	AAx37278
2	2242	99.8	1923	2	AAx37279
3	2183	97.2	1920	2	AAx37277
4	2125.5	94.6	3003	2	AAV82382
5	1986.5	88.4	1299	2	AAQ27516
6	452.5	20.1	1977	2	AAAT8567
7	452.5	20.1	1977	2	AAx05926
8	416.5	18.5	1236	2	AAx05920

9	416.5	18.5	1566	2	AAAT85668
10	416.5	18.5	1962	2	AAAT85695
11	416.5	18.5	1962	2	AAx05929
12	403.5	18.0	1977	2	AAAT85669
13	372	16.6	2121	4	ABL54900
14	346	15.4	2539	2	AAAT61454
15	346	15.4	2809	2	AAAT61455
16	338	15.0	135638	7	ABX34289
17	332.5	14.8	1329	6	ABK74643
18	310.5	13.8	2532	2	AAQ29134
19	307.5	13.7	1859	2	AAAT85677
20	307	13.7	2835	2	AAAT08141
21	307	13.7	4765	2	AAAT08132
22	307	13.7	4765	2	AAAT85670
23	307	13.7	4765	2	AAx05921
24	306.5	13.6	1560	7	ABZ37569
25	306.5	13.6	59816	7	ABZ37516
26	306.5	13.6	59816	7	ABZ37515
27	301.5	13.4	3413	2	AAV72330
28	296	13.2	898	2	AAAT08131
29	290.5	12.9	564	2	AAAT08134
30	290.5	12.9	564	2	AAAT85676
31	283	12.6	3743	9	ADD24901
32	282	12.6	3788	9	ADD24905
33	276	12.3	1306	6	ABL55784
34	276	12.3	1330	6	ABL55787
35	275	12.2	546	4	ABL53453
36	265.5	11.8	2273	2	AAQ04339
37	265.5	11.8	2274	2	AAQ12838
38	265.5	11.8	2274	2	AAQ75859
39	263	11.7	4716	9	ABQ80437
40	263	11.7	4740	7	ABZ58957
41	256.5	11.4	1152	8	ACC85473
42	256.5	11.4	1194	8	ADA32120
43	256.5	11.4	10216	2	AAAT39279
44	256	11.4	1485	6	ABK74647
45	256	11.4	4650	6	ABN71526

ALIGNMENTS

RESULT 1

AAx37278

ID AAx37278 standard; DNA; 1923 BP.

XX	AAx37278;	
AC		
XX		
DT	20-MAR-2003 (revised)	
DT	21-JUL-1993 (first entry)	
XX		
DE	Bacillus alkaline protease encoding DNA.	
XX		
KW	Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;	
KW	washing composition; oxidising agent; ss.	
XX		
OS	Bacillus sp.	
XX		
PN	WO9918218-A1.	
XX		
PD	15-APR-1999.	
XX		
PF	07-OCT-1998; 98WO-JP004528.	
XX		
PR	07-OCT-1997; 97JP-00274570.	
XX		
PA	(KAOS) KAO CORP.	
XX		
PI	Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;	
XX	Shikata S, Nomura M;	
XX		
DR	WPI; 1999-287736/27.	
DR	P-PSDB; AAY17088, AAY17090.	
XX		

PT Alkali protease from Bacillus used in washing powders.

XX
PS Disclosure; Page 58-63; 71pp; Japanese.

XX The invention relates to alkaline proteases produced by strains of
CC Bacillus. The proteases ability to digest casein is not inhibited by
CC oleic acid and they have a high stability to oxidising agents. The
CC alkaline protease of the invention has the following properties: (a) it
CC is active over the pH range 4-13 and has at least 80% of its optimum
CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
CC its ability to digest casein is not inhibited by oleic acid; (e) it has
CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
CC used as enzymes in washing compositions for use in automatic dishwashers
CC and for washing clothes. The stability to oxidising agents allows the
CC enzyme to be an effective component of washing compositions including
CC bleaches. The present sequence represents an alkaline protease encoding
CC DNA. (Updated on 20-MAR-2003 to correct DR field.)

XX SQ Sequence 1923 BP; 576 A; 418 C; 475 G; 454 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,13e-163 Length: 1923
Score: 2247.00 Matches: 434
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-09-985-689A-1 (1-434) x AAX37278 (1-1923)

QY 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
DB 619 AATGATGTTGGCGTGGATTGTCAAAGCGGATGTGGCTCAGACGAGTACGGGTGTAT 678
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 679 GGAACAAGGACAGATCGTCGGTTCGGGATACAGGCTGTGATACAGGTGCGAATGACAGT 738
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 739 TCGATCGATGAGCTTCGCGGGGAAATTAATCTGCATTATATGATTCGGACGCGAAT 798
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
DB 799 AATGCCAATGATACGAATGCTGATGTCATGTCATGTCGCTGGCTGCTGCTTATAGGAAACGCG 858
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
DB 859 TCCACTTAATAAGAAATAGCGCTCAGGCGAATCTAGTCTTCCATCTATCATGATAGC 918
QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
DB 919 GGTGGGGGACTTGGAGGACTTACCTCGAATCTCGAAACCTTATTCAGCCCAAGCATACAGT 978
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyValTyrThr 140
DB 979 GCTGGTGCCGAAATTCATCAAACTCTCGGGAGCGACGATGAATGGGCTTTCACAAACA 1038
QY 141 AspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
DB 1039 GATTCAGAAATGTGGATGACTATGTGCGCAAAAATGATATGACCATCTTTTCGCTGCC 1098
QY 161 GlyAsnGlyGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
DB 1099 GGGAAATGAAGGACCGGAACCGCGGAACCATCAGTGCACCGGACAGCTAAATAATGCAATA 1158
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
DB 1159 ACAGTCGAGCTACGGAAACCTCCGCCCAAGCTTTGGGTCTTATCGGACATATCAAC 1218
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220

DB 1219 CATGTGGCAGCTTCTCTTCACTGGACCGCAAGGATGGAGGATCAACCGGATGTC 1278
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
DB 1279 ATGGCACCGGAAAGTTCATCTATCAGCAAGATCTTCTCTTGCACCGGATTCCTCTTC 1338
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
DB 1339 TGGCGCAACCATGACAGTAAATATGATCATATGGGGAACGTCCTCATGGCTACCGCATC 1398
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
DB 1399 GTTGCTGGAAACGTGGCACAGCTTCGTGAGCATTTGTGAAAAACACAGAGCATCACACCA 1458
QY 281 LysProSerLeuLeuLysAlaAlaIleAlaGlyAlaAspIleGlyLeuGlyTyr 300
DB 1459 AAGCTTCTCTATAAAGCGGCACGTGATTCGGTGGAGCTGACATCCGCTTGGCTAC 1518
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
DB 1519 CCGAACGGTAACCAAGGATGGGACGAGTGACATTGATAAATCCCTGAACGTTGCTAT 1578
QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
DB 1579 GTGACGAGTCCAGTTCCTCTATCCACGACCAAAAGCGACGTACTCGTTTACTGTACT 1638
QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
DB 1639 GCGGCAAGCCTTGAANAATCTCCCTGGTATGCTGATGCCCTCGGAGCACACTGCT 1698
QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
DB 1699 TCGTAACGCTTGTCAATGATCTGGACCTTGTCATTACCGCTCCAAATGGCACAGTAT 1758
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
DB 1759 GTAGAAATGACCTTACTTCCCATACATGATACTGGATGGCGCAATACGTAGAA 1818
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
DB 1819 AATGATATTATTAATGACCAACCAAGCGGACGTATACAATTGAGGTACAGGCTTATAAC 1878
QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
DB 1879 GTACCGTTGGACACAGACCTTCTGTTGGCAATTGTGAAT 1920
RESULT 2
AAX37279
ID AAX37279 standard; DNA, 1923 BP.
XX AAX37279;
AC AAX37279;
XX 20-MAR-2003 (revised)
DT 21-JUL-1999 (first entry)
XX Bacillus alkaline protease encoding DNA.
DE Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
KW washing composition; oxidising agent; ss.
XX Bacillus sp.
OS WO9918218-A1.
PN 15-APR-1999.
XX 07-OCT-1998; 98WO-JP004528.
PF 07-OCT-1997; 97JP-00274570.
XX (KAO) KAO CORP.
XX Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
PI

PI Shikata S, Nomura M;
 XX WPI; 1999-287736/27.
 DR P-PSDB; AAY17091.
 XX
 PT Alkali protease from *Bacillus* used in washing powders.
 XX
 PS Disclosure; Page 63-68; 71pp; Japanese.
 XX
 CC The invention relates to alkaline proteases produced by strains of
 CC *Bacillus*. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
 XX
 SQ Sequence 1923 BP; 578 A; 417 C; 474 G; 454 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2,75e-163 Length: 1923
 Score: 2242.00 Matches: 433
 Percent Similarity: 99.77% Conservative: 0
 Best Local Similarity: 99.77% Mismatches: 1
 Query Match: 99.78% Indels: 0
 DB: 2 Gaps: 0

US-09-985-689A-1 (1-434) x AAX37279 (1-1923)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerTyrGlyLeuTyr 20
 Db 619 AATGATGTTGGGTGGTGAATGTCAGCGGATGTCGTCAGAGCAGCTACGGGTGTAT 678
 Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 679 GGACACGACGATCGTAGCGGTGCGATACAGGCGTTGATACAGCTCGCAATGACAGT 738
 Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 739 TCGATGATGAAGCTTCCCGGGAAATACAGTATATGATGATGCGACGACGAGAT 798
 Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 799 AATGCCAATGATACGATGTCATGTCATGTCATGTCGCTGGCTCCGTATTAGGAAACGGC 858
 Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 859 TCCACTAATAAGGAATGGCGCTCAGCGCAATCTAGTCTTCCAAATCTATCATGAGTACG 918
 Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 919 GGTGGGGAGCTTGAGAGCTACCTTCGAATCTGCAACCTTATTCAGCCAAAGCATACAGT 978
 Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
 Db 979 GCTGGTGCAGAAATTCATACAACTCTCTGGGGAGCAGCAGTGAATGGGGCTTACACAACA 1038
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1039 GATTCCAGAAATGGGATGACTATGTGGCAAAATGATATGACATCTCTTTTCGTGCC 1098
 Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 1099 GCGAATGAAGACCGAAGCGGGCAACCATCAGTCACCGCAGCAGCTAAATAATGCAATA 1158
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200

Db 1159 ACAGTCGGAGCTACGGAAACCTCCGCCCAAGCTTTGGGTCTTATCGGACAATATCAAC 1218
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 1219 CATGTGCACAGTTCTCTTCACGTGGACCCGACAAAGGATGGACGATCAACACCGGATGC 1278
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 1279 ATGCACCGGGAACGTTCACTATCAGCAAGATCTTCTCTTGCCACCGGATTCCTCTTC 1338
 Qy 241 TrpAlaHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
 Db 1339 TGGCGCAACCATGACAGTAAATATGCATACATGGGTGGAACGTCATGGGTACACCGATC 1398
 Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 1399 GTTCTGGAAACGTTGGCACAGCTTCGTGAGCATTTTGTGAAACACAGGCAATCACACCA 1458
 Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaIleGlyLeuGlyTyr 300
 Db 1459 AAGCCTTCTCTATTAAAGCGGCACCTGATTGCGCGTGCAGCTGACATCGGCTTGGGTAC 1518
 Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 1519 CCGAACGTAACCAAGGATGGGACGAGTGCATGATGATTAATCCCTGACGTTGCCAT 1578
 Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 1579 GTGAACGAGTCCAGTTCTCTATCCACAGCCAAAGCGACGCTACTCGTTTACTGTACT 1638
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 Db 1639 GCCGCAAGCCTTTGAAATCTCCCTGGTATGGTCTGATGCCCTGGAGCAACACTGCT 1698
 Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1699 TCCGTAAACGCTTGTCAATGATCTCGACCTGTCTATTACCGCTCCAAATGGCACAGTAT 1758
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 Db 1759 GTAGAAATGACTTTACTTCGCCATCAATGATAACTAAGTGGGATGGCCCAATAACGTAGAA 1818
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1819 AATGATTTTATTATGCACCAACAAAGCGGACGATATACAATTGAAGTACAGGCTTATAAC 1878
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1879 GTACCGTGGACCAACAGAACTTCTCGTTGGCAATTGTGAAT 1920

RESULT 3
 AAX37277
 ID AAX37277 standard; DNA; 1920 BP.
 XX
 AC AAX37277;
 XX
 XX 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 DE *Bacillus* alkaline protease encoding DNA.
 XX
 KW Alkaline protease; *Bacillus*; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent; ss.
 XX
 OS *Bacillus* sp.
 XX
 PN WO9918218-A1.
 XX
 PD 15-APR-1999.
 XX
 PF 07-OCT-1998; 98WO-JP004528.
 XX

PR 07-OCT-1997; 97JP-00274570.
XX (KAOS) KAO CORP.
PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
PI Shikata S, Nomura M;
XX WPI; 1999-287736/27.
DR P-PSDB; AAV17087, AAV17089.
XX Alkali protease from Bacillus used in washing powders.
XX Disclosure; Page 53-58; 71pp; Japanese.
CC The invention relates to alkaline proteases produced by strains of
CC Bacillus. The proteases ability to digest casein is not inhibited by
CC oleic acid and they have a high stability to oxidising agents. The
CC alkaline protease of the invention has the following properties: (a) it
CC is active over the pH range 4-13 and has at least 80% of its optimum
CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
CC its ability to digest casein is not inhibited by oleic acid; (e) it has
CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
CC used as enzymes in washing compositions for use in automatic dishwashers
CC and for washing clothes. The stability to oxidising agents allows the
CC enzyme to be an effective component of washing compositions including
CC bleaches. The present sequence represents an alkaline protease encoding
CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
XX
SQ Sequence 1920 BP; 576 A; 411 C; 473 G; 460 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 9.57e-159 Length: 1920
Score: 2183.00 Matches: 418
Percent Similarity: 99.31% Conservative: 13
Best Local Similarity: 96.31% Mismatches: 3
Query Match: 97.15% Indels: 0
DB: 2 Gaps: 0

US-09-985-689A-1 (1-434) x AMX37277 (1-1920)

QY 1 AsnAspValAlaArgGlyIleValIysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 616 AATGATGTGCCAGAGGATTTGTCAAGCGGATGTGGCACAGCAGCTACGGTTTGTAT 675
QY 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGACAAAGGCCAGATTCTCCAGTTGCCGATCTGGATTGGATACAGGAAGAAGACAGT 735
QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATGCATGACGCTTCGGCGGTAAATACAGCACTATATGCACTGGGTCCGACGAAT 795
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 796 AATGCGAATGATACGAACGGTCTATGTTACCCATGTCGAGTTCGGTATTAGGAATGGC 855
QY 81 SerThrAsnIscGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 856 GCACAGATTAAGAGATGGCCCTCAAGCGAATCTGGTTTTCAATCCATCATGATAGC 915
QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 916 AGTGGTGGGCTTGAGGCTTCCCTCCATCTGCAAAACCTTATTTCAGCCCAAGCAATTCAGT 975
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyValAlaValAsnGlyValTyrThrThr 140
Db 976 GCAGGTGCCAGATTTATACAAACTCTGGGGGGCAGCGGTGATGGGCCCTACAGACA 1035
QY 141 AspSerArgAsnValAspAspTyrValArgIysAsnAspMetThrIleLeuPheAlaAla 160
Db 1036 GATTCCAGAAATGTGGATGACTATGTAAGGAAATGATATGACGATCTTTTCGGCGCT 1095

QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaIysAsnAlaIle 180
Db 1096 GGGAAATGAAGCGCGGACCGGATCCATCAGTCAGCTGGTACGGCTAAACGCGCAT 1155
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1156 ACAGTCGGCGCACACCGAAACCTGCGTCCAAGCTTCGGTTCCTATGATGATATATTAAC 1215
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyValArgIleIysProAspVal 220
Db 1216 CAGTTGACAGTTCTCTCCCGTGGCCGACAAAGATGGCGAATCAAGCTGATGTC 1275
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 1276 ATGGCGCCAGGACATACATTTTATCAGCAAGATCTCTCTGCAACCGATTCCTCTCTTC 1335
QY 241 TrpAlaAsnHisAspSerIysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 1336 TGGCGCAATCATGACGACAAATATGCTATATGGGTGGACGCTCCATGGCAACCGATT 1395
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIysAsnArgGlyIleThrPro 280
Db 1396 GTTGGCGGGAATGTCACAGCTCCCGTGGATCTTTGTGAAAAATAGAGGAATCACTCT 1455
QY 281 LysProSerLeuLeuIysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 1456 AAGCTTTCCTATTGAAGCAGCTTTGATTGCGAGTGTCTGATGATTTGGATTGGGTTAT 1515
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 1516 CCGAAGCAAAACCAAGATGGGCGCGAGTGACCTGGATAAATCGTTGAACGTTGCTAT 1575
QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnIysAlaThrTyrSerPheThrAlaThr 340
Db 1576 GTGAACGAATCCAGTGCCTTATCACTAGCCAAAAGGACATATACCTTTACTGCAACG 1635
QY 341 AlaGlyLysProLeuIysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 1636 GCGGCGCAGCCATTGAAAATCTCCCTGGTATGTCGATGGCCCTGCAAGCACTACTGCT 1695
QY 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 1696 TCTGTAAACCTGTGCAATGATTGGATTGGTTCATTACAGCCAAACGAAACAGATAT 1755
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
Db 1756 GTCGGGAATGACTCTCAGCACCATTTGCAATAACTGGGATGGCGCGCAATACGTAGAA 1815
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1816 AATGTATTTTAAATTCGCCCCAAAGTGAACATATACCATTTGAGGTGCAAGCATATAAT 1875
QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1876 GTCCGGTTGGACCAAAAACCTCTCGTTGGCAATTTGTGAAC 1917

RESULT 4
AAV82382 standard; DNA; 3003 BP.
ID AAV82382
XX
AC AAV82382;
XX
DT 12-APR-1999 (first entry)
XX
DE Bacillus JP170 protease gene.
XX
KW Protease; detergent; surfactant; leather processing; debittering;
KW flavour; ss.
XX
OS Bacillus sp.
XX
FH Key
CDS 846..2771

FT /*tag= a
 FT sig_peptide 846..944
 FT /*tag= b
 FT mat_peptide 1470..2768
 FT /*tag= c

XX WO9856927-A2.

XX 17-DEC-1998.

XX 09-JUN-1998; 98WO-US012005.

XX 12-JUN-1997; 97US-00873479.

XX (NOVO) NOVO NORDISK BIOTECH INC.

XX Sloma A, Christianson L;

XX WPI; 1999-080908/07.

XX P-PSDB; AAW89547.

XX Novel protease from *Bacillus subtilis* LC20 - useful in laundry and
 PT dishwashing detergents and for leather processing.

XX Claim 11; Page 52-53; 77pp; English.

XX This nucleotide sequence encodes a novel protease (see AAW82382) of
 CC *Bacillus* sp. JP170 (NCIB 12513). The sequence in plasmid p170BAN is
 CC contained in *Bacillus subtilis* LC20 NRRL B-21680. The protease gene was
 CC isolated from chromosomal DNA of JP170 following preparation of probes
 CC based on protease N-terminal and internal peptides (see AAW89549-50).
 CC screening of chromosomal libraries, isolation of the 3' end of the gene
 CC by inverse PCR (see AAW84410-11), reconstruction of 5' and 3' ends and
 CC PCR amplification (see AAW82412-16). Claimed recombinant host cells can
 CC be used in a method for producing the protease. The protease is used in
 CC laundry and dishwashing detergents, for institutional and industrial
 CC cleaning, and for leather processing, as well as for debittering and
 CC enhancing the degree of hydrolysis of protein hydrolysates, for flavour
 CC development through hydrolysis of proteins, degradation of undesired
 CC peptides and in enzymatic synthesis of peptides. It has enhanced
 CC stability towards oxidation under alkaline conditions, e.g. towards
 CC bleaching agents of the peroxy type. The invention also provides mutant
 CC cells in which the protease activity is diminished. Such cells can be
 CC used for the production of heterologous recombinant proteins

SQ Sequence 3003 BP; 982 A; 504 C; 645 G; 872 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 4,35e-154 Length: 3003
 Score: 2125.50 Matches: 406
 Percent Similarity: 97.93% Conservative: 19
 Best Local Similarity: 93.55% Mismatches: 8
 Query Match: 94.59% Indels: 1
 DB: 2 Gaps: 1

US-09-985-689A-1 (1-434) x AAW82382 (1-3003)

Qy 1 AsnAspValAlaArgGlyValLeuValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 1470 AATGACGTGGCCGTGGCATTTGGTGAAGCAGACGTCGCACAAATAACITTTGCTTATAT 1529
 Qy 21 GlyGlnGlyGlnLeuValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 1530 GGACAAGGACAGATTGTAGCAGTTGCTGATCTGGGCTTGATACAGAAAGAAATGACAGT 1589
 Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 1590 TCGATGTCATGAGCATTTCCCGGCTAGATTACCGCACTATATGCATCTGGGCGACAGCAAT 1649
 Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 1650 AACGCCAATGATCCAAATGGACATGGAAACCATTTCTGCTGGATCTGTGTAGGAAT--- 1706

RESULT 5

Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 1707 GCTACAAATAAAGGATGGCAGCGCAATCTAGTCTCTTCAATCTATTATGGATAGT 1766
 Qy 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 1767 GGTGGAGGGCTGGAGGACTACTCTGTAATCTACAAACATTATTTCAGTCAAGCATATAGT 1826
 Qy 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAsnGlyAlaTyrThrThr 140
 Db 1827 GCTGGAGCGAGATTTCATACGAATTCATGGGGGGCTCCAGTAAACGGTGCCTATACGACA 1886
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1887 GACTCTCGAAATGTTGATGATTATGTGAGAAATAATGATATGACGATCTCTTTTTCGGGCC 1946
 Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 1947 GGAAATGAGGGACCCAGCTAGCGGTACAATCAGTGCACCGAGAACAGCAAAATATGCGATT 2006
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 2007 ACAGTTGGGCGACCCGAAACCTACGTCCAGCTTCGGATCTTATGCGGATATATTTAAC 2066
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 2067 CATGTTGCTCAATTCCTTCACGAGGTCCTACTAGAGATGACGTATTAAAGCGGACGTC 2126
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 2127 ATGGCACCCAGGTAGCTATATCTCTCTGCTAGATCATCTTAGCTCCAGATTCCTCATTC 2186
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProile 260
 Db 2197 TGGCAAAACCATGATAGTAATATGCTACATGGGTGCTACTTATATGGCTACTCCATT 2246
 Qy 251 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 2247 GTAGCAGGTAATGTTGCACAAATTAAGGAGCATTTTGTGAAAAATAGAGGGGTAACTCCT 2306
 Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 2307 AAGCTTCCCTTTTAAAGCTGCTTTTAAATGTCAGTCTCTCGGATGTTGGACTTGGCTTT 2366
 Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 2367 CCAATGTTAAACCAAGGATGGGAAGAGTAACTAGATAAATCCCTAAATGTGCGATT 2426
 Qy 321 ValAsnGluSerSerSerLeuSerThrSerClnLysValAlaThrTyrSerPheThrAlaThr 340
 Db 2427 GTGAATGAACGAGCGCCCTTTATCAACAGTCAAAAGCAACATATTCGTTTACGGCTCAA 2486
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 Db 2487 GCTGGTAAACCTTTAAAAATATCACTTTGTTGGTCAGATGCACCGGTAGCAGCGCA 2546
 Qy 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGluTyr 380
 Db 2547 TCACCTAACTTTAGTGAATGATTTAGACTTAGTAATCATCGCCAAATGGAACTAAATAC 2606
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 Db 2607 GTCGGAATAGCTTTACAGCACCGTATGATAACAATTTGGGATGCGCAGAACCAACGCGAA 2666
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 2667 AATGTGTTTATCAATGCTTCCTCAAGCGGAACGTATACAGTCGAAGTGCAGGCTTACAAT 2726
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 2727 GTACCAATAGTCCCGAAACCTTTTCTTTAGCGATTGTACAT 2768

424

AAQ27516
 ID AAQ27516 standard; DNA; 1299 BP.
 XX
 AC
 AAQ27516;
 XX
 DT 05-FEB-1993 (first entry)
 XX
 DE Alkali-protease Ya enzyme gene.
 XX
 KW Alkali resistance; surface active agent resistance; detergent improver;
 KW ss.
 XX
 OS Bacillus sp. Y.
 XX
 FH Key Location/Qualifiers
 CDS 1..1299
 FT /*tag= a
 FT
 XX JP04197182-A.
 XX
 PD 16-JUL-1992.
 XX
 PF 28-NOV-1990; 90JP-00327110.
 XX
 PR 28-NOV-1990; 90JP-00327110.
 XX
 PA (LIOY) LION CORP.
 XX
 WI; 1992-288440/35.
 DR P-PSDB; AAR26274.
 DR
 XX
 PT DNA coding alkali-protease Ya enzyme - has good alkali and surfactant
 PT resistance and improves detergency.
 PS Claim 3; Page 2; 17pp; Japanese.
 XX
 CC The sequence is that of the alkali-protease Ya enzyme gene which can be
 CC used in the recombinant production of Ya enzyme. Ya enzyme is excellent
 CC in alkali resistance and surface active agent resistance and improves
 CC detergency
 XX
 SQ Sequence 1299 BP; 401 A; 246 C; 297 G; 355 T; 0 U; 0 Other;
 XX

Alignment Scores:
 Pred. No.: 8,12e-144 Length: 1299
 Score: 1986.50 Matches: 379
 Percent Similarity: 94.01% Conservative: 29
 Best Local Similarity: 87.33% Mismatches: 25
 Query Match: 88.41% Indels: 1
 DB: Gaps: 1

US-09-985-689A-1 (1-434) x AAQ27516 (1-1299)

Qy 1 AsnAspValAlaArgGlyLeuValAlaAspValAlaGlnSerSerThrGlyLeuTyr 20
 Db 1 AATGATGTAGCAAGAGGATAGTAAAGCTGATGTGCACAAACAAATACCGATTATAT 60
 Qy 21 GlyGlnGlyClnLeuValAlaValAlaAspThrGlyLeuAspThrGlyArgAspSer 40
 Db 61 GGACAGGTCACTAGTTCAGTAGCGGACACAGGCTTAGACAGTCTGTAACGATAGT 120
 Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 121 TCTATGATGAACATTCGCGGAAATACACAGCTCTTTACGCGTTAGGAAGAACTAAT 180
 Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 181 AATGCGAGTATCCGAATGGCATGGCAGACACATGTAGCAGGTCTGTACTTCGTAAT--- 237
 Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 238 GCCTTAAATAAAGGAATGGTCCGCAAGCTAACTTAGTCTTCCAAATCTATTATGATAGC 297

Qy 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 298 AGCGGAGGATTAGTGGCTTACATCGAATCTTAATACGTTATTATGATGAGTTGGAAT 357
 Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
 Db 358 GCTGGAGCAAGAATTCACTAACTCTTGGGAGGCCCAAGTAAATGGAGCGTACACTGCT 417
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 418 AACTCGAGCAAGTGGATGATGATGTTTGAATAATGATATGATGCGGTACTTTTTCGAGCT 477
 Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 478 GGTAATGAAGTCTTAATCAGGAACAATTAGTGTCCAGGTACACGCAAAATGCTATT 537
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 538 ACGTCGCGCGCAACGGAACAACTATCCCAAGCTTCGGTTCGATAGCAGTAACCCCAAT 597
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 598 CATATTGCACAATTTTCATCGAGAGGAGCTACGAGGATGGACGAATTAAGCTGACGTA 657
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 658 ACAGCTCCTGGAACATTTATTATCAGCACGTTCTTCTTAGCTCCAGACTCTTCGTT 717
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db 718 TGGCGCAATTAACAGTAATAACGCGTATATGGCGGTACCTCCATGCGCACACCTATT 777
 Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 778 GTTCAGGGAATCTCGCAATTAACGTAATTTTATATAAATAAGAGGTATTACTCCT 837
 Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 838 AAGCTTCTTTAATAAAGCTGACATTATCGCTGGTGCTACTGATCTTGGTTAGGATAT 897
 Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 898 CCTAGTGTGACCAAGCTGGGCGCTGTACTCTAGATAAATCGTTAAATGTAGCGTAT 957
 Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 958 GTCAATGAGCAACATGATTAGCCACAGACAAAGAACACGTTATTCGTCCAAGCACAA 1017
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 Db 1018 GCGGGTAAACCTTTAAATACTCGTTAGTATGACAGATGCTCTCGGAAAGTACAACTGCA 1077
 Qy 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1078 TCTTATACATAGTAAATGATTAGATCTAGTTATTACTGCTCCGATGACAAATAATAT 1137
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 Db 1138 GTAGAAATGATTCTAGTTATCTTATGATAATAACTGGATGCTCGCAACATGTTGAG 1197
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1198 AACGTATTTTAAACGCTCGCAATCTGGCACTATATATTAATTGAGGTTCAGCGTATAT 1257
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1258 GTACCATCTGGCCACAGCGTTTCTCACTAGCTATCGTACAT 1299

RESULT 6
 AAT85667
 ID AAT85667 standard; DNA; 1977 BP.
 XX
 AC AAT85667;

XX 17-OCT-2003 (revised)
 DT 20-APR-1998 (first entry)
 XX
 DE Thermococcus protease coding sequence.
 XX
 KW Protease; research reagent; thermal stability; thermococcus celer; ss.
 XX
 OS Thermococcus celer; DSM-2476.
 XX
 SS WO9721823-A1.
 PN
 XX
 FN 19-JUN-1997.
 PD
 XX
 XX 07-NOV-1996; 96WO-JP003253.
 XX
 XX 12-DEC-1995; 95JP-00323285.
 PR
 XX
 XX (TAKI) TAKARA SHUZO CO LTD.
 PA
 XX Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
 PI Tsunawawa S, Kato I;
 PI
 XX WPI; 1997-332794/30.
 DR
 XX P-PSDB; AAW24121.
 XX
 XX Protease(s) and genes encoding them obtained from Thermococcus and
 PT Pyrococcus strains - have extremely high thermal stability and are useful
 PT industrially and as research reagents.
 XX
 XX Claim 3; Page 86-87; 159pp; Japanese.
 PS
 XX This sequence represents the coding sequence for the protease from
 CC Thermococcus celer DSM-2476. This sequence encodes a protease of the
 CC invention. The proteases of the invention have extremely high thermal
 CC stability. The proteases can be used as research reagents, and
 CC industrially in the food, drug and chemical industries. (Updated on 17-
 CC OCT-2003 to standardise OS field)
 XX
 SQ Sequence 1977 BP; 453 A; 659 C; 554 G; 311 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1,728-25 Length: 1977
 Score: 452.50 Matches: 138
 Percent Similarity: 44.66% Conservative: 67
 Best Local Similarity: 30.07% Mismatches: 153
 Query Match: 20.14% Indels: 101
 DB: 2 Gaps: 18

US-09-985-689A-1 (1-434) x AAT85667 (1-1977)

QY 8 ValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnLeuValAla 27
 Db
 433 ATAGGGGCGGATACCGTCTGGAACTCCCTCGGCTACGACGGAAGCGGTGGTGGTGGC 492
 QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
 Db
 493 ATCTGCGATACGGGTATAGACGCGAAC-----CACCCGATCTGAG 534
 QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
 Db
 535 GCGAAGGTATAGGGTGGTATACGACGCGCTCAACGCGCAGGTGACCCCTACGATGACCCAG 594
 QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys----- 84
 Db
 595 GGACACGGAAACCCAGCTTCGGGGTATCGTTGCCGGAACCGGCGGTAACTCCAGTAC 654
 QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 101
 Db
 655 ATAGGGGTGCGCCCGCGCGAAGTCTGTCGCGCTCAAGGTCTCGTGGCGCGACGCTTCG 714
 QY 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 121

RESULT 7
 AAX05926
 ID AAX05926 standard; DNA; 1977 BP.

Db 715 GGAAGCGTCTCACCATCATCGCGGGTGTGACTGGTCTCCAGAACAGACAAAGTAC 774
 QY 122 GlyAlaArgIle-----HisThrAsnSer 129
 Db 775 GGGATAAGGGTCAACAACCTCTCCCTCGGCTCTCCAGAGCTCCGACGGAACCACTCC 834
 QY 130 TrpGlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrVal 149
 Db 835 CTCAGTCAGGCGGTCAACAACGCTGGGACGCC----- 867
 QY 150 ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyThr 169
 Db 868 -----GGTATAGTAGTCTCGCTCGCGCGCAACAGCGGCGCCGACACCTACACC 918
 QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
 Db 919 GTGCGTCAACCGCGCGGAGCAAGGTCAATACCGTGGTGA----- 963
 QY 190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
 Db 964 -----GTTGACAGCAACGACACATCGCCAGCTTCTCCAGCAGGGA 1005
 QY 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229
 Db 1006 CCGACCGCGACGGAAGGCTCAAGCGGAGTCTGTCGCCCGCGGCTTGACATCATAGCC 1065
 QY 230 AlaArgSerSerLeuAlaProAspSerSerPheThrAlaAsnHisAspSerLysTyrAla 249
 Db 1066 CCGCGCGCCAGC-----GGAACCAAGCATGGCGACCCCGCATAAACGACTACTACAC 1116
 QY 250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly---AsnValAlaGlnLeu 268
 Db 1117 AAGGCTCTGGAAACAGCATGGCCACCGCGACGTTTCGGGCGTGGCGGCTCATCTC 1176
 QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeuLys 286
 Db 1177 CAGGCGCCAC-----CCGAGTGGACCCCGACAAAGTGAG 1212
 QY 287 AlaAlaLeuIleAlaGlyAla-----AlaAspIleGlyLeu 298
 Db 1213 ACGGCTCTCATGACACCGCGCATAGTCGCCCAAGGAGATAGCGGACATCGCTTAC 1272
 QY 299 GlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeu----- 316
 Db 1273 GGTGGC-----GGTAGGTGAAGTCTCAAGGCCATCAAGTAC 1311
 QY 317 ---AsnValAlaTyrValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyr 335
 Db 1312 GACGACTACGCCAAGCTCACCTTCACCGCTCCGTCGCGCAAGGAGCGCCACCCAC 1371
 QY 336 SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaPro 355
 Db 1372 ACCTTCGAGTCAGCGCGCCACCTTCGTGACCGCCACCTCTACTGGAC----- 1422
 QY 356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAspLeuValIleThrAlaPro 375
 Db 1423 -----ACGGCTCGAGCGACATCGACTCTTACTCTACGACCCC 1461
 QY 376 AsnGlyThrGlnTyrValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGly 395
 Db 1462 AAGCGGAACGAG---GTTGACTACTCTACACCGCTACTAC----- 1500
 QY 396 ArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGlu 415
 Db 1501 -----GGCTTCGAGAAGGTGCGCTACTACAACCGCGCGAACCCTGGACGCTCAAG 1554
 QY 416 ValGlnAlaTyrAsnValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1555 GTCTGTCAGCTACAAG-----GGCGCGGGAACACTACAGTGCACGTCTGTCAGC 1602

QY 416 ValcAlaTyrAsnValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 DB 1555 GTGCTCAGTACAAAG-----GGCGCGCAACTACCAAGCTCAGCTCGTCAGC 1602

RESULT 8

AA05920
 ID AAX05920 standard; DNA; 1236 BP.

XX AC AAX05920;

XX DT 06-MAY-1999 (first entry)

XX DE Hyperthermostable protease fragment encoding DNA.

XX KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 XX XW additive; drug; washing agent; foodstuff; chemical synthesis; ds.

XX OS Pyrococcus furiosus.

XX PN WO9856926-A1.

XX PD 17-DEC-1998.

XX PF 04-JUN-1998; 98WO-JP002465.

XX PR 10-JUN-1997; 97JP-00151969.

XX PA (TAKI) TAKARA SHUZO CO LTD.

XX PI Takakura H, Morishita M, Shimojo T, Asada K, Kato I;

XX DR WPI; 1999-080907/07.

XX DE P-PSDB; AAW94836.

PT Recombinant hyperthermostable protease from Pyrococcus furiosus - and
 PT gene encoding it, for large scale production of the protease for
 PT industrial use.

PS Claim 6; Page 37-38; 82pp; Japanese.

XX The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially Pyrococcus furiosus). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at
 CC 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Gly-Asn-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially bacillus strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable for
 CC industrial use, can be used as an additive for drugs, washing agents and
 CC foodstuffs and for chemical synthesis

SQ Sequence 1236 BP; 368 A; 251 C; 309 G; 308 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 5.85e-23 Length: 1236
 Score: 416.50 Matches: 141
 Percent Similarity: 42.55% Conservative: 59
 Best Local Similarity: 30.00% Mismatches: 148
 Query Match: 18.54% Indels: 122
 Gaps: 20

US-09-985-689A-1 (1-434) x AAX05920 (1-1236)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24

DB 24 GTCTGCAGCTCAAGTATATGGCACTTACGTTTGGAACTTGGATATGATGTTCTGGAAT 83

QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44

DB 84 CACAATAGGAATAATATGACACTGGAATTGAC-----GCTTCTCACC 125

QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
 DB 126 AGATCTCCAGGAAGAAGTA-----ATTGGGTGGGTAGATTTTGTCAATGG 170
 QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
 DB 171 TAGGAGTTATCCATACGATGACCATGACATGGAACCTCATGTAGCTTCAATAGCAGCTGG 230
 QY 78 YAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
 DB 231 TACTGGAGCAGCAAGTAATGCAAGTACAGGGAAATGGCTCCAGGAGTAAGCTGGCGGG 290
 QY 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
 DB 291 AATTAGGTTCTAGTGCCGATGTTCTTGGAGCATATCTACTATATAATTAAGGAGTGA 350
 QY 112 mThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
 DB 351 GTGGCCCGTTGATAACAAGATAAGTACGGAATTAAAGTCAATTAACTCTTCTCTGGTTC 410
 QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
 DB 411 AAGCCAGAGCTCAGATGCTACTAGCTCTAAGTCAGCTCTTAAATGCAGCGTGGATGC 470
 QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
 DB 471 T-----GGATTAGTTGTTGGTTCGCCG 494
 QY 160 sGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI1 180
 DB 495 TGGAAACAGTGGACCTTAACAGTATACAATGGTTCTCCAGCAGCTGCAGCAAGTTAT 554
 QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
 DB 555 TACAGTTGGAGCC-----GTTGACACAGTATGA 581
 QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
 DB 582 TGTATAACAAGCTTCTCAAGCAGAGGGCCAACTCCAGCAGCGAGCTTAAGCCCTGAGGT 641
 QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
 DB 642 TGTGCTCCAGAAACTGGATAATGTCTGCCAGACCAAGT-----GGAACATGACAT 692
 QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProI1 260
 DB 693 GGCTCAACCAATAATGACTATTACACAGCAGCTCTCTGGACATCAATGGCAACTCTCA 752
 QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
 DB 753 CGTAGCTGTATGTCAGCCCTCTGCTCAA-----GCACACCC 791
 QY 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
 DB 792 GAGCTGGACTCCAGACAAAAGTAAAAACAGCCCTCATAGAAACTGCTGATATCGTAAAGCC 851
 QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
 DB 852 AGATGAATAGCCGATATAGCTACGTTGCA-----GGTAGGGT 890
 QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330
 DB 891 TAATGCATACAAGGCTATAAAC-----TACGATAACTATGCAAGAGTAGTCTTCACTGG 944
 QY 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
 DB 945 ATATGTTGGCAACAAGCAGCAAACTCCAGCTCTGTTATAGCGAGCTTCGTTGCT 1004
 QY 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
 DB 1005 AACTGCCACATATATCTGGACAATGCCAAT----- 1035
 QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385

```

Db 1036 -AGCGACCTTGATCTTTACCTCAGCATCCCAATGGAACCCAG---GTTGACTACTCTTTA 1091
QY 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1092 CACGGCTACTAT-----GGATTGCAAAAGGTTGGTTATTA 1127
QY 405 nAlaProGlnSerGlyThrThrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1128 CAACCCCAACTGATGGAAACATGGACAATTAAAGTTGTGAAGCTACAGC-----GGAAG 1178
QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1179 TGCAAACTATCAAGTAGATGTGTAAGT 1206

RESULT 9
AAT85668
ID AAT85668 standard; DNA: 1566 BP.
XX
AC AAT85668;
XX
DT 17-OCT-2003 (revised)
DT 20-APR-1998 (first entry)
XX
DE Pyrococcus furiosus protease coding sequence.
XX
KW Protease; research reagent; thermal stability; pyrococcus furiosus; ss.
XX
OS Pyrococcus furiosus; DSM-3638.
XX
FH Key Location/Qualifiers
FT CDS 1..1566
FT /tag= a
FT /transl_except= (pos: 1282..1284, aa: Xaa)
FT /note= "Xaa= Gly, Val"
XX
FN WO921823-A1.
XX
PD 19-JUN-1997.
XX
PF 07-NOV-1996; 96WO-JF003253.
XX
PR 12-DEC-1995; 95JP-00323285.
XX
PA (TAKI ) TAKARA SHUZO CO LTD.
XX
PI Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
PI Tsunashawa S, Kato I;
XX
DR WPI; 1997-332794/30.
DR P-PSDB; RAW24122.
XX
PT Protease(s) and genes encoding them obtained from Thermococcus and
PT Pyrococcus strains - have extremely high thermal stability and are useful
PT industrially and as research reagents.
XX
PS Claim 7; Page 90-91; 159pp; Japanese.
XX
CC This sequence represents the coding sequence for the protease from
CC Pyrococcus furiosus DSM-3638. This sequence encodes a protease of the
CC invention. The proteases of the invention have extremely high thermal
CC stability. The proteases can be used as research reagents, and
CC industrially in the food, drug and chemical industries. (Updated on 17-
CC OCT-2003 to standardise OS field)
XX
SQ Sequencé 1566 BP; 467 A; 342 C; 372 G; 384 T; 0 U; 1 Other;

Alignment Scores:
Pred. NO.: 7.74e-23 Length: 1566
Score: 416.50 Matches: 141
Percent Similarity: 42.55% Conservative: 59
Best Local Similarity: 30.00% Mismatches: 148
Query Match: 18.54% Indels: 122

```

```

DB: 2 Gaps: 20
US-09-985-689A-1 (1-434) x AAT85668 (1-1566)
QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
Db 24 GTCTGCAGCTCAAGTTATGGCAACTTACGTTTGGAACTTGGGATATCATGTTCTGGAAT 83
QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisG1 44
Db 84 CACAATAGGAATAATTGACACTGGAAATTGAC-----GTTCTCATCC 125
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAlaAsn-- 63
Db 126 AGATCTCCAGGAAAAAGTA-----ATTGGTGGGTAGATTGTTGTCATGG 170
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 171 TAGGAGTTATCCATACGATGACCATGACATGGAACCTCATGTAGCTTCAATAGCAGCTGG 230
QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnValPh 94
Db 231 TACTGGAGCAAGTAATGCAAGTACAAGGGGAATGCTCCAGGAGCTAAGCTGGCGGG 290
QY 94 eClnSerIleMet-----AspSerGlyGlyGlyLeuGlyLeuProSerAsnLeuG1 112
Db 291 AATTAAAGTTCTAGTGCCTGCTCTGGAGCATATCTACTATAATTAAGGAGGTGA 350
QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTyrGlyVal 132
Db 351 GTGGGCCCTTGATTAACAAGATAAGTACGGAATTAAGTTCATTATCTTCTCTGTTGTC 410
QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
Db 411 AAGCAGAGCTCAGATGCTGACTGACGCTCTAAGTCAGGCTGTTAATGCAGCGTGGGATGC 470
QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 471 T-----GGATTAGTTGTTGTTGCTGCCG 494
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI1 180
Db 495 TGGAAACAGTGGACCTTACAAGTATACATCGTTCTCCAGCAGCTCCAGCAAGTAT 554
QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 555 TACAGTTGGAGCC-----GTTGACAAAGTATGA 581
QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 582 TGTTATAACAAGCTTCTCAAGCAGAGGGCCCAACTGCAGACGGCAGGCTTAAGCTGAGGT 641
QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 642 TGTGTCTCCAGGAACCTGGATAATTGCTGCCAGAGCAAGT-----GGAACCTAGCAT 692
QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProI1 260
Db 693 GGGTCAACCAATTAATGACTATTACACAGCAGCTCCTGGGACATCAATGGCACTCTCTCA 752
QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 753 CGTACTGGTATGTCAGGCCCTCTTGCTCCAA-----GCACACCC 791
QY 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 792 GAGCTGGACTCCAGACAAAGAAAAACAGCCCTCATAGAAAACCTGCTGATATCGTAAAGCC 851
QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
Db 852 AGATGAATAGCCGATATAGCTACGGTGCA-----GATAGGGT 890
QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330

```

Db 891 TAATGCATACAGGCTATAAAC-----TACGATACTATGCAAGCTAGTGTCTACTGG 944
Qy 330 rGlnLyAala-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
Db 945 ATATGTTGCCAACAAAGCAGCCAACTCACCAGTTCGTTATTAGCGGACCTCGTTCGT 1004
Qy 345 uLysLeSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1005 AACTGCCACATTATACGGGCAATGCCAAT----- 1035
Qy 365 lAsnAspLeuAspLeuValThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
Db 1036 -AGGACCTTGATCTTTACCTCTACGATCCCAATGGAAACAG---GTTGACTACTCTTA 1091
Qy 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheLeAs 405
Db 1092 CACGCCCTACTAT-----GGATTCCAAAAGGTTGGTTATTA 1127
Qy 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1128 CAACCCAACTATGGAACATGGCAATTAAGTTGTAGCTACAGC-----GGAAG 1178
Qy 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1179 TGCAAACTATCAAGTAGATGTGGTAAGT 1206
RESULT 10
AAT85695
ID AAT85695 standard; DNA; 1962 BP.
AC AAT85695;
XX 17-OCT-2003 (revised)
DT 20-APR-1998 (first entry)
XX Pyrococcus furiosus PFUS protease coding sequence.
DE Pyrococcus furiosus PFUS protease coding sequence.
XX Protease; research reagent; thermal stability; pyrococcus furiosus; ss.
XX Pyrococcus furiosus; DSM-3638.
XX WO9721823-A1.
XX 19-JUN-1997.
XX 07-NOV-1996; 96WO-JP003253.
XX 12-DEC-1995; 95JP-00323285.
XX (TAKI) TAKARA SHUZO CO LTD.
XX Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
PI Tsunawawa S, Kato I;
XX WPI: 1997-332794/30.
DR P-PSDS; AAN24129.
XX Protease(s) and genes encoding them obtained from Thermococcus and
PT Pyrococcus strains - have extremely high thermal stability and are useful
PT industrially and as research reagents.
XX Disclosure; Page 123-125; 159pp; Japanese.
XX This sequence represents the coding sequence for the protease from
CC Pyrococcus furiosus DSM-3638. This sequence encodes a protease of the
CC invention. The proteases of the invention have extremely high thermal
CC stability. The proteases can be used as research reagents, and
CC industrially in the food, drug and chemical industries. (Updated on 17-
CC OCT-2003 to standardise OS field)
XX Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.01e-22 Length: 1962
Score: 416.50 Matches: 141
Percent Similarity: 42.55% Conservative: 59
Best Local Similarity: 30.00% Mismatches: 148
Query Match: 18.54% Indels: 122
DB: 2 Gaps: 20
US-09-985-689A-1 (1-434) x AAT85695 (1-1962)
Qy 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGly 24
Db 420 GTCTGCAGCTCAAGTTATGGCAACTTACGTTGGACTGGGATATGATGTTCTGGAAAT 479
Qy 24 nLleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG 44
Db 480 CACAATAGGAATAATTGACACTGGAATTGAC-----GTTCTCATCC 521
Qy 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAGGAAGAATGTA-----ATTGGTGGGTAGATTTTGTCAATGG 566
Qy 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG 78
Db 567 TAGGAGTTATCCATACGATGACCATGGACATGGAACATCATGTAGCTTCAATAGCAGCTG 626
Qy 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGCAAGTACAGGGAATGCTCCAGAGCTAAGCTGGCGG 686
Qy 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyLeuGlyLeuProSerAsnLeuG 112
Db 687 AATTAAGGTTCTAGTGCCGATGTTCTGGAAGCATATCTACTATAATTAAGGAGTTGA 746
Qy 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 747 GTGGCCCTTCATACAAAGAAATAGTACGGAATTAAGTCAATATCTTCTCTGTGTTTC 806
Qy 132 a-----AlaValAsnGlyAlaTyrThrTh 140
Db 807 AAGCCAGAGCTCAGATGGTACTGACGCTTAAGTCAGCTGTTAATGCAGCGTGGGATGC 866
Qy 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 867 T-----GGATTAGTTGTGTGTTGCCG 890
Qy 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI 180
Db 891 TGGAAACAGTGGACCTTAAAGTATACATCGTTCTCCAGCAGCTGCAGCAAGTTAT 950
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 951 TACAGTTGGAGCC-----GTTGACAAAGTATGA 977
Qy 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 978 TGTATAAACAAGCTTCTCAAGCAGAGGGCCAACTGCAGACGCGAGGTTAAGCTTAGGT 1037
Qy 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 1038 TGTGTCTCCAGAAACTGGATAATTGCTGCCAGCAAGT-----GGAACCTAGCAT 1088
Qy 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIl 260
Db 1089 GGGTCAACCAATTAATGACTATTACACAGCAGCTCTCTGGGACATCAATGCAACTCTCA 1148
Qy 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 1149 CGTAGCTGTATTGACGCCCTCTTGTCTCCAA-----GCACACCC 1187
Qy 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 1188 GAGCTGGACTCCAGACAAAGTAAAGAACAGCCCTCATAGAAACTGCTGATATCGTAAAGCC 1247

QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyValGva 310
 Db 1248 AGATGAATAGCGGATATAGCTACGGTCCA-----GGTAGGGT 1286
 QY 310 ThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330
 Db 1287 TAATGCATCAAGGCTATAAAC-----TACGATAACTATGCAAGCTAGTGTTCACCTGG 1340
 QY 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
 Db 1341 ATATGTTGCCACAAAGGCGAGCAAACTCCACGATCGTTATTAGCGGAGCTTCGTCGT 1400
 QY 345 uLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
 Db 1401 AACTGCGCATATATCTGGGCAATGCCAAT----- 1431
 QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
 Db 1432 -AGCGACCTTGATCTTACTCTACGATCCCAATGGAACCAAG---GTGACTACTCTTA 1487
 QY 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
 Db 1488 CACCGCCTACTAT-----GGATTGGAAGGTTGGTTATTA 1523
 QY 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
 Db 1524 CAACCCAACTGATGGAACTGACAAATTAAGTTGTAGCTACAGC-----GGAAG 1574
 QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
 Db 1575 TGCAAACTATCAAGTAGATGTGTAAGT 1602

RESULT 11

AA05929
 ID AA05929 standard; DNA; 1962 BP.

XX
 AC AAX05929;
 XX
 DT 06-MAY-1999 (first entry)
 XX
 DE Hyperthermostable protease encoding DNA.
 XX
 KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.
 XX
 OS Pyrococcus furiosus.
 XX
 PN WO9856926-A1.
 XX
 PD 17-DEC-1998.
 XX
 PF 04-JUN-1998; 98WO-JP002465.
 PR 10-JUN-1997; 97JP-00151969.
 XX
 PR (TAKI) TAKARA SHUZO CO LTD.
 PA
 XX Takakura H, Morishita M, Shimojo T, Asada K, Kato I;
 XX WPI; 1999-080907/07.
 XX P-PSDB; AAW94841.
 XX
 PT Recombinant hyperthermostable protease from Pyrococcus furiosus - and
 PT gene encoding it, for large scale production of the protease for
 PT industrial use.
 XX
 PS Disclosure; Page 59-60; 82pp; Japanese.
 XX
 CC The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially Pyrococcus furiosus). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at

CC 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Gly-Asn-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially Bacillus strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable for
 CC industrial use, can be used as an additive for drugs, washing agents and
 CC foodstuffs and for chemical synthesis
 XX
 SQ Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,01e-22 Length: 1962
 Score: 416.50 Matches: 141
 Percent Similarity: 42.55% Conservative: 59
 Best Local Similarity: 30.00% Mismatches: 148
 Query Match: 18.54% Indels: 122
 DB: 2 Gaps: 20

US-09-985-689A-1 (1-434) x AA05929 (1-1962)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
 Db 420 GTCTGCAGCTCAAGTTATGCAACTTACGTTTGGAACTTGGGATATGATGTTCTCGAAT 479
 QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMethISG1 44
 Db 480 CACATAGGAAATATGACACTGGAATTGAC-----GCTTCTCATCC 521
 QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
 Db 522 AGATCTCCAAGGAAAGTA-----ATTGGTGGGTAGATTTCCTCAATGG 566
 QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
 Db 567 TAGGAGTTATCCATACATGACCATGACATGGAAGTCTATGACCTTCATAGCAGCTGG 626
 QY 78 YAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
 Db 627 TACTGGAGCAGCAAGTAATGGCAAGTACAAGGGAATGGCTCCAGGAGCTAAGCTGCGGG 686
 QY 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
 Db 687 AATTAAGGTTCTAGTGCCGATGGTCTGTGAAGCATATCTACTATATTAATTAAGGGAGTTGA 746
 QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
 Db 747 GTGGCGCGTTGATAACAAAGATAAGTAGCGAATTAAGGTCATTAACTCTTCTTGGTTC 806
 QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
 Db 807 AAGCCAGAGCTCAGATGGTACTGACGCTCTAAGCTCAGGCTGTTAATGCGAGCTGGATGC 866
 QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
 Db 867 T-----GGATTAGTTGTTGGTGGCGG 890
 QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIl 180
 Db 891 TGGAAACAGTGGACCTAACAAAGTATACAATCGGTTCTCCAGCAGCTGCAAGCAAGTTAT 950
 QY 180 eThrValGlyValaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
 Db 951 TACAGTTGGAGCC-----GTTGACCAAGTATGA 977
 QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
 Db 978 TGTATAACAAGCTTCTCAAGCAGACGGCCAACTGACAGACGGCAGGCTTAAGCTGAGGT 1037
 QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
 Db 1038 TGTTCTCCAGGAACATGGGATAATTGCTGCAGACGAAGT-----GGAAGTACAT 1088


```

Db 1038 CGTCGCCCGCGGCTTGACATCATAGCCCGCGCCAGC-----GGAACACGACAT 1088
QY 240 eTPAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProI 260
Db 1089 GGGCACCCCGATTAACGACTACTACCAAGGCTCTGGAACAGCATGGCCGCCCGCA 1148
QY 260 eValAlaGly---AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLeu 279
Db 1149 CGTTTCGGCGGCTTGGCGGCTCATCTCCAGGCCAC----- 1185
QY 279 rProLys-----ProSerLeuLeuLysAlaLeuLeuAlaGlyAla----- 293
Db 1186 -CGAGCTGACCCCGACAGGTGAAGACCGCCCTCATCGAGCCGCGCATAGTCGC 1244
QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyAr 309
Db 1245 CCCCAAGGAGATAGCGGACATCGCTACGGTGG-----GGTAG 1283
QY 309 gValThrLeuAspLysSerLeu-----AsnValAlaTyrValAsnGluSerSerSe 326
Db 1284 GGTGAACGCTCTACAGGCCATCAAGTACGACGACTACGCCAAGCTCACCTCACCGGCTC 1343
QY 326 rLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThrAlaGlyLysProLeu 346
Db 1344 CGTCGCCGACAGGGAAGCGCCACCCACACTTCGACGTCAGCGGCCACCTTCGTGAC 1403
QY 346 sIleSerLeuValTyrSerAspAlaProAlaSerThrThrAlaSerValThrLeuVal 366
Db 1404 CGCACCCCTCTACGGGAC-----ACGGGCTCGAG 1433
QY 366 nAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPhe 386
Db 1434 CGACATCGACCTCTACCTTACGACCCCAACGGAACGAG-----GTTGACTACTCCTAC 1490
QY 386 rSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAl 406
Db 1491 CGCTACTAC-----GGCTTCGAGAAGGTCGGTACTACAA 1526
QY 406 aProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPro 426
Db 1527 CCGGACCGCGGACCTGGACGCTCAAGTCTCAGCTACAG-----GGCGCGC 1577

RESULT 13
ABL54900
ID ABL54900 standard; DNA; 2121 BP.
AC ABL54900;
XX
XX 11-SEP-2003 (revised)
DT 31-MAY-2002 (first entry)
XX
DE T. yonsei subtilisin-like serine protease coding sequence.
XX
KW Subtilisin-like serine protease; ss.
XX
XX Thermoanaerobacter yonseiensis.
XX
XX Key Location/Qualifiers
XX CDS 142..179
XX FT /*tag= a
XX FT /product= "subtilisin-like serine protease"
XX
XX KR2000072141-A.
XX
XX PN
XX
XX PD
XX
XX PF
XX
XX 04-AUG-2000; 2000KR-00045411.
XX

```

```

PR 04-AUG-2000; 2000KR-00045411.
XX
XX (KIMY/) KIM Y S.
XX
XX Chang HJ, Kim DH, Byun YR, Kim YS;
XX
XX WPI; 2001-298092/31.
XX
XX P-ESDB; ABB09483.
XX
XX New DNA sequence of thermophilic protein decomposition enzyme and protein
XX derived therefrom.
XX
XX Claim 1; Page 6; 15pp; Korean.
XX
XX This sequence represents the DNA encoding the Thermoanaerobacter yonsei
XX subtilisin-like serine protease of the invention. (Updated on 11-SEP-2003
XX to standardise OS field)
XX
XX SQ Sequence 2121 BP; 712 A; 410 C; 425 G; 554 T; 0 U; 20 Other;

Alignment Scores:
Pred. No.: 2,96e-19 Length: 2121
Score: 372.00 Matches: 130
Percent Similarity: 46.29% Conservative: 51
Best Local Similarity: 33.25% Mismatches: 114
Query Match: 16.56% Indels: 96
DB: 4 Gaps: 19

US-09-985-689A-1 (1-434) x ABL54900 (1-2121)
QY 6 GlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnIle 25
Db 445 GGAATCACAAAA-----GCACGGAGTGATTTTGGAGTCACAGCAAAAATATAACA 495
QY 26 ValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAla 45
Db 496 ATAGCAATTATTGACACAGGTATAGCGAAATCACGTTGACCTCTCA----- 543
QY 46 PheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn----- 63
Db 544 ---CGTGAAAAATA-----ATAGGATGGAAGACATTATCAACAACAAA 585
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 586 ACTACACATACGACGACAAATGCCATGGAACTCAGTAGCAAGTATTGCTGCAGGTACA 645
QY 80 GlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSer 96
Db 646 GGTGCTGGAACAGCTCTTTACAAAGCGTTGCTCCTGATGCTTTGTTGGTGAATAAAA 705
QY 97 IleMetAspSerGlyGly-----LeuGlyGlyLeuProSerAsn 110
Db 706 GTTTTAGATGCAATGGAATGGAAGCGGACGATGAGCAGCTGTAACGAGGAATTGACTGGGCT 765
QY 111 LeuGln-----ThrLeuPheSerGlnAla 118
Db 766 GTTCAAAATAAAGATGTATACGGAAATCAAGTTATAAATTAGCTTCGCATCTCTACA 825
QY 119 TyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyr 138
Db 826 AGTTCTGATGGA-----ACTGACTCTACCTCATTAGCAGTGAAT----- 864
QY 139 ThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPhe 158
Db 865 -----ACGCGATTGTAGTAGTTGTA 897
QY 159 AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsn 178
Db 898 GCACGAGAAACCTCTGGCCCTGCACAAATACACCATAGGTCGCCCTGTGTGCGGAAAAA 957
QY 179 AlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsn 198
Db 958 GCCATAACAGTCGACGAATGGCAGATGTA-----GGTGAACCTTGGCTTTAAC 1005

```

QY 199 IleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysPro 218
 ::
 Db 1006 CTT-----GCAAGCTTTTCAGCGCGGTCTACTCTGCTGACGGAAGAATAAACCT 1056
 QY 219 AspValMetAlaProGlyThrPheIleLeuSerAlaArgSerLeuAlaProAspSer 238
 ::
 Db 1057 GACATTGGCGCCAGGATATAATAATAACTGCCCAAG----- 1095
 QY 239 SerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThr 258
 ::
 Db 1096 -----GCAATTCTGTAAATGGATATGTAACATACAGCGGTACACATGGCAACA 1146
 QY 259 ProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIle 278
 ::
 Db 1147 CTTTTGTAGCAGAACTGTTGCTCTTATGCTTAAC-----GCTAATCCAAATCTC 1197
 QY 279 ThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeu 298
 ::
 Db 1198 ACTCCAATGATGCA-----AAAAATATAATATGCTACTCCAAAAAGCTGGGGCCT 1251
 QY 299 GlyTyrProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnVal 318
 ::
 Db 1252 CCAAGCAAAACGTTACTATGTTGTCAGGAAGATGGCTATGAAGCTATAAGGGTA 1311
 QY 319 Ala-----TyrValAsnGluSer 324
 ::
 Db 1312 GCAGGTAATTTAGAGAAATAATATTGATGTACCAAAATCATTTATATAT-----TCA 1365
 QY 325 SerSerLeuSerThrSerGlnLysAla---ThrTyrSerPheThrAlaThr---AlaGly 342
 ::
 Db 1366 GGTATCTGCCGGTTCGTGTACAGTGATCTGGACTTTTAAACGCAAAATACAAAGT 1425
 QY 343 LysProLeuLysIleSerLeuValTrpSerAsp 353
 ::
 Db 1426 TATCCAATCGCAATAACATTAATAATTCCTGAC 1458
 RESULT 14
 AAT61454
 ID AAT61454 standard; DNA; 2539 BP.
 XX AC AAT61454;
 XX DT 06-OCT-1997 (first entry)
 XX DE Streptomyces viridosporus dhpa gene.
 XX KW asymmetric hydrolase, dhpa; 4-substituted-1,4-dihydropyridine;
 KW derivative; Streptomyces viridosporus; ester; chiral; synthesis;
 KW cardiovascular; treatment; hypertension; ischaemic heart disease; ds.
 XX OS Streptomyces viridosporus.
 XX FH Key Location/Qualifiers
 XX CDS 338..2539
 XX FT /*tag= a
 XX FT /note= "no stop codon given"
 XX FT 950..2509
 XX FT /*tag= b
 XX FT /note= "encodes AAW13666"
 XX PN WO9705243-A1.
 XX PD 13-FEB-1997.
 XX PF 30-JUL-1996; 96WO-JP002147.
 XX PR 31-JUL-1995; 95JP-00212975.
 XX PR 29-FEB-1996; 96JP-00067478.
 XX PA (SAOC) MERCIAN CORP.
 XX PI Arisawa A, Matsufuji M, Tsuruta T, Dobashi K, Nakashima T;

PI Isshiki K, Yoshioka T;
 XX WPI: 1997-145582/13.
 DR P-PSDB; AAW13666, AAW13667.
 XX
 PT Asymmetric hydrolase gene derived from Streptomyces viridosporus - acts
 on 4-substituted-1,4-dihydropyridine derivatives to produce chiral
 derivatives useful for synthesis of cardiovascular drugs.
 XX Claim 3; Page 49-55; 78pp; Japanese.
 CC This sequence is the Streptomyces viridosporus dhpa gene which encodes an
 asymmetric hydrolase which acts on 4-substituted-1,4-dihydropyridine
 derivatives. The enzyme allows the efficient conversion of 4-substituted-
 1,4-dihydropyridine esters to chiral partially hydrolysed derivatives,
 for use in the synthesis of cardiovascular drugs suitable for the
 treatment of e.g. hypertension and ischaemic heart disease
 XX Sequence 2539 BP; 433 A; 959 C; 867 G; 280 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 3,688-17 Length: 2539
 Score: 346.00 Matches: 129
 Percent Similarity: 42.00% Conserved: 52
 Best Local Similarity: 29.93% Mismatches: 167
 Query Match: 15.40% Indels: 84
 DB: 2 Gaps: 13
 US-09-985-689A-1 (1-434) x AAT61454 (1-2539)
 QY 2 AspValAlaArgGlyLeuValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
 Db 953 GACAGCTGCTGGGAGATCGGCGCCCAAGGCTGTCGCGC---GGTACGACGCG 1009
 QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 41
 Db 1010 AAGGCGGTGAAGATCGCGTCTCTGACACCGGTGTCACACGAGC----- 1054
 QY 42 MetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
 Db 1055 ---CATCGGACCTGAAGGCGCGGTGACCGCTCCAAAGACTTCACGCGCGCGCGCC 1111
 QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
 Db 1112 GCGCGCGACAGGTGGCGCACCGCACCCACGCTCGCTCGATCGCGCGGCGCGCGCC 1171
 QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
 Db 1172 CAGTCCAAAGGCAAGTACAAGGCGGTGCGACCGCGCGCGGATCCTCAACGCAAGTC 1231
 QY 98 MetAspSerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
 Db 1232 CTCGACGACTCGGT-----TTCGGCGACGACTCCGCGCATCCTCGCGGATGAGTGG 1285
 QY 118 AlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAlaValAsnGlyAl 137
 Db 1286 GCGCGCGCGAGCGCGCGAGCTGCTCAACATGAGCTGGCGCGGATCGACACACCGGAG 1345
 QY 137 aTyrThrThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLe 157
 Db 1346 ACCGACCGCTGGAGCGCGCG-GTGCAACAAGTCCCGCGAGAGGCGTCTGTTTCG 1404
 QY 157 uPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaL 177
 Db 1405 CATCGCGCGCGCAACGAGGCGCGGAG-----TCGATCGTTCGCGCGCGCGCGGA 1458
 QY 177 sAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAs 197
 Db 1459 CGCCGCGCTTCACCGTCGCGCGC-----GTGCA 1485
 QY 197 pAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLysAspGlyArgI 216
 Db 1486 CGACAAGGACAAAGCTCGCGGACTTCTCTCCACGCGCGCGCGCTCGCGCGCGCGCAT 1545

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 15, 2004, 23:21:18 ; Search time 84 Seconds
(without alignments)

2867.246 Million cell updates/sec

Title: US-09-985-689A-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAVNPVGPQTFLAIVN 434

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DBV=xlh
-Q=/cgn2_1/USPTO.spool/US0985689/runat_10032004_112904_19757/app_query.fasta_1.583
-DB=issued Patents NA -QMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOFCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0985689@cgn1_1_44@runat_10032004_112904_19757 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WASN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2247	100.0	1923	4	US-09-509-814A-5
2	2242	99.8	1923	4	US-09-509-814A-7
3	2183	97.2	1920	4	US-09-509-814A-3
4	2125.5	94.6	3003	2	US-08-873-479-41
5	452.5	20.1	1977	3	US-08-894-818B-2
6	452.5	20.1	1977	3	US-09-445-472-11
7	416.5	18.5	1236	4	US-09-445-472-2
8	416.5	18.5	1566	3	US-08-894-818B-4
9	416.5	18.5	1962	3	US-08-894-818B-34
10	416.5	18.5	1962	4	US-09-445-472-15
11	403.5	18.0	1977	3	US-08-894-818B-6
12	351	15.6	2539	3	US-09-000-016-3

13	351	15.6	2539	4	US-09-514-340-3
14	346	15.4	2809	3	US-09-000-016-1
15	346	15.4	2809	4	US-09-514-340-1
16	310.5	13.8	2532	1	US-07-671-376C-4
17	307.5	13.7	1859	3	US-08-894-818B-15
18	307	13.7	2835	1	US-08-750-532-2
19	307	13.7	4765	1	US-08-750-532-8
20	307	13.7	4765	3	US-08-894-818B-7
21	307	13.7	4765	4	US-09-445-472-5
22	296	13.2	898	1	US-08-750-532-7
23	290.5	12.9	564	1	US-08-750-532-11
24	290.5	12.9	564	3	US-08-894-818B-14
25	276	12.3	1306	4	US-09-966-921A-1
26	276	12.3	1330	4	US-09-966-921A-5
27	256.5	11.4	1194	4	US-09-328-352-3407
28	256.5	11.4	10216	2	US-08-875-154-1
29	252	11.2	1628	4	US-09-634-238-212
30	251	11.2	840	1	US-08-434-255-7
31	251	11.2	840	1	US-08-459-967-7
32	251	11.2	840	1	US-08-460-327-7
33	251	11.2	840	1	US-08-459-871-7
34	251	11.2	840	3	US-09-024-532-1
35	251	11.2	840	4	US-09-104-623A-1
36	251	11.2	840	4	US-09-019-532-1
37	251	11.2	840	4	US-09-417-355A-1
38	251	11.2	840	4	US-09-705-185-1
39	251	11.2	840	4	US-09-338-746-1
40	251	11.2	1110	1	US-08-434-255-5
41	251	11.2	1110	1	US-08-459-967-5
42	251	11.2	1110	1	US-08-460-327-5
43	251	11.2	1110	1	US-08-459-871-5
44	251	11.2	1191	1	US-08-434-255-3
45	251	11.2	1191	1	US-08-459-967-3

ALIGNMENTS

RESULT 1
US-09-509-814A-5
; Sequence 5, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAKKI, KATSUHIISA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1923)
US-09-509-814A-5

Alignment Scores:
Pred. No.: 1.14e-224 Length: 1923
Score: 2247.00 Matches: 434
Percent Similarity: 100.00% Conservative: 0

```

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-09-985-689A-1 (1-434) x US-09-509-814A-5 (1-1923)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
DB 619 AATGATGTTGCGCGTGGAAATTTGCAAGCGGATGGCTCAGACAGCTACGGGTTGTAT 678
QY 21 GlyGlnGlyGlnIleValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 679 GGCAAGGACAGATCGTACGGTTCGCGATACAGGCTTGATACAGGTGCGAATGACAGT 738
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 739 TCGATCATGAACCTTCGCGGGAAATTAATGCAATATATGCAATGGAGCGGACGAT 798
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
DB 799 AATGCCAATGATACGAATGGTATGTTGATGCTACGGCTTGATGCTGGCTGCGTATTAGGAAACGCG 858
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
DB 859 TCCACTAATAAGGATGGCGCTCAGGCGAATCTAGTCTTCATCTATCATGATGATGAC 918
QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
DB 919 GGTGGGGGAGCTTGGAGGACTTACCTCGAATCTCGAAACCTTATTGAGCCAGCATACAGT 978
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyValaTyrThr 140
DB 979 GCTGTGCGCAGAAATTCATCAAACTCTGGGAGGACGAGTGAATGGGCTTACCAACA 1038
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
DB 1039 GATTCAGAAATGTGATGACTATGTGCGCAAAATGATATGACGATCTTTTCGCTGCC 1098
QY 161 GlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
DB 1099 GGGAAATGAAGGACCGGACCGGAGCCATCATGTCAGCACAGGACAGCTAAATAATGCAATA 1158
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
DB 1159 ACAGTCGGAGTACGAAACCTCGCCCAAGCTTTGGTCTTATGCGGACATATCAAC 1218
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
DB 1219 CATGTGGCACAGTCTCTTCACGTGGACCGCAAAAGGATGGACGGATCAAAACCGGATGTC 1278
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
DB 1279 ATGCCACCGGAAACGTTCTACTATACAGCAAGTCTTCTTGCACCGGATTCCTCTTC 1338
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
DB 1339 TGGCGGAACCATGACAGTAAATATGACATCATGGTGGAAACGTCCTCATGGCTACACCGATC 1398
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
DB 1399 GTTGCTGGAAACGTGGCAAGCTTCGTGAGCATTTTGTGAAACACAGGCGATCAACCA 1458
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
DB 1459 AAGCCTTCTCTATTAAGGCGCACTGATTCGCGGTGCACTGACATCGGCTTGGCTAC 1518
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValaTyr 320
DB 1519 CCGAAGGTAACCAAGATGGGACGAGTGACATTGGATAAATCCCTGAAACGTTGCCCTAT 1578
QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
DB 1579 GTCAACGAGTCCAGTCTCTATCCACACCGCAAAAGGACGCTACTCGTTTACTGCTACT 1638

RESULT 2
US-09-509-814A-7
; Sequence 7, Application US/09509814A
; Patent No. 6378227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSURISHA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUMI
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1923)
US-09-509-814A-7

Alignment Scores:
Pred. No.: 3,8e-224 Length: 1923
Score: 2242.00 Matches: 433
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 1
Query Match: 99.78% Indels: 0
DB: 4 Gaps: 0

US-09-985-689A-1 (1-434) x US-09-509-814A-7 (1-1923)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
DB 619 AATGATGTTGCGCGTGGAAATTTGCAAGCGGATGGCTCAGACAGCTACGGGTTGTAT 678
QY 21 GlyGlnGlyGlnIleValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 679 GGCAAGGACAGATCGTACGGTTCGCGATACAGGCTTGATACAGGTGCGAATGACAGT 738
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60

```


Db 976 GCAGGTGCCAGAAATTCATACAACTCTGGGGGGCAGCGGTGAATGGGGCTCACAGACA 1035
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1036 GATTCAGAGAAATGTGGATGACTATATAGGAAAAATGATACAGATTCTCTTTGGCGCT 1095
 Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 1096 GGGATGAAAGCCGACGCGGTACCATCAGTGACCTGGTACCGCTAAAGACGCCATA 1155
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 1156 ACAGTCGGCGCACCGCAACCTGGCTGCAAGCTTCGGTTCTATGCGAGATAAATTAAAC 1215
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyValGlyIleLysProAspVal 220
 Db 1216 CACGTTGCACAGTTCTCTCCGTCGCGCCGACAAAGATGGCGAATCAAGCCTGATGTC 1275
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 1276 ATGGCGCCAGGACATACATTTTATCAGCAAGATCTTCTTGCCACCGGATTCCTCTTC 1335
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
 Db 1336 TGGCGAATCATGACAGCAAAATATGCTATATGGTGGAACTGCTGATGGTAAACACCGATT 1395
 Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 1396 GTTGGGGGAATGTTGCACAGCTCCGTGAGCATTTTGTGMAAAATAGAGGAATCACTCCT 1455
 Qy 281 LysProSerLeuLeuLysAlaIleAlaIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 1456 AAGCCTTCCTTATGAAACAGCTTTGATTGAGGTGCTGCTGATGTTGGATTTGGTTAT 1515
 Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 1516 CCGAACGGAAACCAAGATGGGGCGAGTGACCTGGATAAATCGTTGAACGTTGCCAT 1575
 Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 1576 GTGAACGAATCCAGTSCCTATCAACTAGCCAAAAGCGACATATACCTTTACTGCAACG 1635
 Qy 341 AlaGlyProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 Db 1636 GCGGGCAACCAATGAAATCTCCCTGGTATGGTGGATGCCCTGCGAGCACTACTGCT 1695
 Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1696 TCTGTAAACCTGCTCAATGATTGGATTTGGTCAITACAGCACCAACCGAACAAGATAT 1755
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyValArgAsnValGlu 400
 Db 1756 GTGGGAATGACTTCTCAGCACCATTTGACATTAAGTGGATGGCGGCGCATTAACGTAGAA 1815
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1816 AATGTATTATTAAATTCGCCCAAGTGAACATATACCATTTGAGGTGCAAGCATATAAT 1875
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1876 GTGCCGTTGGACCAAAACCTTCTCGTTGGCAATTTGTAAC 1917

RESULT 4

US-08-873-479-41
 ; Sequence 41, Application US/08873479
 ; Patent No. 5891701
 ; GENERAL INFORMATION:
 ; APPLICANT: Sionia, Alan
 ; APPLICANT: Lynne, Christanson
 ; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
 ; TITLE OF INVENTION: Having Protease Activity
 ; NUMBER OF SEQUENCES: 57
 ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5891701o No. 5891701disk of No. 5891701th America
 ; STREET: 405 Lexington Avenue
 ; CITY: New York
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10174
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/873,479
 ; FILING DATE: 12-JUN-1997
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Agria, Cheryl H
 ; REGISTRATION NUMBER: 34,086
 ; REFERENCE/DOCKET NUMBER: 5251.000-US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-867-0123
 ; TELEFAX: 212-878-9655
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 41:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3003 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA
 ; US-08-873-479-41

Alignment Scores:
 Pred. No.: 1,12e-211 Length: 3003
 Score: 2125.50 Matches: 406
 Percent Similarity: 97.93% Conservative: 19
 Best Local Similarity: 93.55% Mismatches: 8
 Query Match: 94.59% Indels: 1
 DB: 2 Gaps: 1

US-09-985-689A-1 (1-434) x US-08-873-479-41 (1-3003)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 1470 AATGACGTGCGCGTGGCATTTGAAAGCAGACGTCGACAAATAACCTTGGCTTATAT 1529
 Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 1530 GGACAAAGACAGATTGTAGCAGTTGCTGATCTGGGCTTGATACAGAAAGAAATGACAGT 1589
 Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 1590 TCGATGATGAGCAATTCGCGGTAAAGATTACCGCACTATATGCACTGGGACAGCAAT 1649
 Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 1650 AACGCCAATGATCAAAATGGACATGGAAACCCATGTTGCTGGATCTGTGTAGGAAT--- 1706
 Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 1707 GCTACAAATAAAGGATGGCAGCCGCAAGCAATCTAGTCTTTCAATCTATTATGATAGT 1766
 Qy 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 1767 GGTGAGGCGCTGGGAGGACTACCTGCTAATCTCAAAACATTATTTCAGTCAAGCATATAGT 1826
 Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyValAlaValAsnGlyAlaTyrThr 140
 Db 1827 GCTGAGCGAGAAATTCATGAATTAATGAGGGGCTCCAGTAAACGGTGTCTATACGACA 1886
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1887 GACTCTCGAAATGTTGATGATTATGTGAGAAAAAATGATATGACGATCTCTTTTGGGCC 1946

161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1947 GGAATGAGGACGAGTACGGTACATCATCGTCCAGGACGACGAAAAATGCGATT 2006
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 2007 ACAGTTGGGGCAACCGAAAAACCTACGTCGAAGCTTCGGATCTTATCGGATATATTAAC 2066
Qy 201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyValGlyIleLysProAspVal 220
Db 2067 CATGTTCTCAATTCCTTCCAGAGGCTCTACAGAGTGGAGTATTAAGCCGACGTC 2126
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 2127 ATGGCACCAGTACGTATATCTCTCTGTAGATCATCATCTAGCTCCAGATTCCTCATTC 2186
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 2187 TGGCAAAACCATGATGATGATAATATGCTACATGGGTGTTACTTCTATGGCTACTCCAAT 2246
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 2247 GTACAGGTAATGTGACAAATTAAGGAGCATTTTGTGAAAATAGAGGGGTAACTCCT 2306
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 2307 AAGCCTTCCCTTTTAAAGCTGCTTTAATTTGAGGTGCTCGGATGTGGACTTGGCTTT 2366
Qy 301 ProAsnGlyAsnGlnGlyTTPGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 2367 CCAATGGTACCAAGAGTGGGAGAGTAACTGTAGTAAATCCCTAAATGTCCGATTT 2426
Qy 321 ValAsnGluSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 2427 GTGAATGAACAGAGCCCTTTTCAACAAAGTCAAAAGCAACATATTCGTTACGGTCAA 2486
Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 2487 GCTGGTAACCTTTAAATAATATCACTTTGTTGTCAGATGCACAGTAGCAGCGGA 2546
Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 2547 TCATAACTTTAGTGAATGATTAGACTTAGTAACTCACTGCACCAAAATGAACATAATAC 2606
Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
Db 2607 GTCCGAAATGACTTTACAGCACCGGTATGATAACAATTTGGGATGGCAGAAACACGTCGA 2666
Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 2667 AATGTGTTTATCAATGCTCCTCAAGCGGAACGTATACATCGAAGTGCAGGCTTACAT 2726
Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 2727 GTACCAATATTCGCAACACCTTTCTTTTACGATTGTACAT 2768

RESULT 5

US-08-894-818B-2
; Sequence 2, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masenori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Kunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark

STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/894,818B
APPLICATION NUMBER: 08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1977 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-818B-2

Alignment Scores:

Pred. No.: 2,57e-37 Length: 1977
Score: 452.50 Matches: 138
Percent Similarity: 44.66% Conservative: 67
Best Local Similarity: 30.07% Mismatches: 153
Query Match: 20.14% Indels: 101
DB: 3 Gaps: 18

US-09-985-689A-1 (1-434) X US-08-894-818B-2 (1-1977)

Qy 8 ValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnIleValAla 27
Db 433 ATAGGGGGCCGATACCGTCTGGAACTCCCTCGGTACGACGGAACGGGTGGTGGTGGC 492
Qy 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 493 ATCTGTCGATACGGGTATAGACGCGAAC-----CACCCCGATCTGAAG 534
Qy 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
Db 535 GCGAAGGTTCATAGGTGCTGATACGCGCGCTCAACGCGAGTCCGCCCTCTAGCATGACCAG 594
Qy 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys----- 84
Db 595 GGACACGGAAACCCAGTTCGGGTATCGTTGCCGGAACCGGCGCGGTAACTCCAGTAC 654
Qy 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 101
Db 655 ATAGCGTCCGCGCGCGCGGCGGAAAGTCTGTCGGCTCAAGGTTCCTCGGTCCGCGCGTTCG 714
Qy 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 121
Db 715 GGAACGCTTCACCATCATTCGCGGTGTGACTGGGTCTGTCGAGAACAGAACAGTAC 774
Qy 122 GlyAlaArgIle-----HisThrAsnSer 129
Db 775 GGGATAAGGGTCATCAACCTCTCCCTCGGTCTCTCCAGAGCTCCGACGGAACCGACTCC 834

```

QY 130 TrpGlyAlaValAsnGlyAlaTyrThrAspSerArgAsnValAspAspTyrVal 149
DB 835 CTCAGTCAGCGCGTCAACAGCGCTGGGACGCC----- 867
QY 150 ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyThr 169
DB 868 -----GGTATAGTCTGCGTCCGCGCGCAACAGCGCGCGCAACACCTACACCC 918
QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
DB 919 GTGGCTACCCCGCCCGCGAGCAAGTCATACCGTCGGTGCA----- 963
QY 190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
DB 964 -----GTTGACAGCAACGACACATCGCCAGCTTCTCCAGCGAGGGA 1005
QY 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229
DB 1006 CCAGCCGCGAGCGAAGGCTCAAGCGGAAGTGTGCGCCCGCGGTGACATCATAGCC 1065
QY 230 AlaArgSerSerLeuAlaProAspSerSerPheTyrAlaAsnHisAspSerLysTyrAla 249
DB 1066 CCGCGCGCCAGC-----GGAACGACGATGGGCAACCCCGATAAACGACTACTACACC 1116
QY 250 TyrMetGlyThrSerMetAlaThrProIleValAlaGly-----AsnValAlaGlnLeu 268
DB 1117 AAGCGCTCGAACCAGCATGGCCACCCCGCATGTTTCGGGCGTGGCGCGCTCATCTC 1176
QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeuLys 286
DB 1177 CAGGCCAC-----CCGAGTGGACCCCGGCAACAGGTGAAG 1212
QY 287 AlaAlaLeuIleAlaGlyAla-----AlaAspIleGlyLeu 298
DB 1213 ACCGCGCTCATCAGACCGCGGACATAGTCCGCCCAAGAGATAGCGGACATCGCTAC 1272
QY 299 GlyTyrProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeu----- 316
DB 1273 GGTGGC-----GGTAGGTGAACGTCTACAAAGGCCATCAAGTAC 1311
QY 317 ---AsnValAlaTyrValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyr 335
DB 1312 GAGGACTACCCCAAGCTCACTTCAACGGGTCCGTCGCGCAACAGGGGAAGCGCCACCC 1371
QY 336 SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValThrSerAspAlaPro 355
DB 1372 ACCTTCAGCTAGCGCGCCACCTTCGAGCCGCCACCTCTACTGGGAC----- 1422
QY 356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAspValIleThrAlaPro 375
DB 1423 -----ACGGGCTCGAGCGACATGACCTCTACCTCTACGACCC 1461
QY 376 AsnGlyThrGlnTyrValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGly 395
DB 1462 AACGGGAACGAG---GTGACTACTCTACCGCCCTACTAC----- 1500
QY 396 ArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrThrIleGlu 415
DB 1501 -----GGCTTCGAGAAGGTGCGTACTACAACCCCGACCGCGGAACCTTGACGGTCAAG 1554
QY 416 ValGlnAlaTyrAsnValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
DB 1555 GTCGTACGTACAG-----GGCGGCGGAACCTACACAGGTGACGCTCGTCAGC 1602

```

RESULT 6

```

US-09-445-472-11
; Sequence 11, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko

```

```

; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1377
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-11

```

Alignment Scores:

```

Pred. No.: 2 57e-37 Length: 1977
Score: 452.50 Matches: 136
Percent Similarity: 44.66% Conservative: 67
Best Local Similarity: 30.07% Mismatches: 153
Query Match: 20.14% Indels: 101
DB: 4 Gaps: 18

```

US-09-985-689A-1 (1-434) x US-09-445-472-11 (1-1977)

```

QY 8 ValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnIleValAla 27
DB 433 ATAGGGCCGATACCGTCTGGAACCTCCCTCGGTACGACGGAAGCGGTGGTGGTGGC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
DB 493 ATCTGTCATAGGTATAGACCGAAC-----CACCCCGATCTGAAG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
DB 535 GGCAAGGTCATAGCTGGTACGACCGCGTCAACGGCAGGTGCGACCCCTACGATGACCAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys----- 84
DB 595 GGACACGGAACCCACGTTGCGGTATCGTTCGCGGAACCGGACGCTTAATCTCCAGTAC 654
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 101
DB 655 ATAGCGTTCGCGCCCGCGCGAAGCTCGTCGGGTCAAGGTTCTCGGTGCGGACGGTTCG 714
QY 102 GlyLysLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 121
DB 715 GGAAGCGTCTCCACCATCATCGCGGTGTTGACTGGGTCTCCAGAACAAAGGCAAGTAC 774
QY 122 GlyAlaArgIle-----HisThrAsnSer 129
DB 775 GGATATAGGTCAACACCTCTCCCTCGGTCTCCCGAGAGCTCCGACGGAACCGACTCC 834
QY 130 TrpGlyAlaAlaValAsnGlyAlaTyrThrAspSerArgAsnValAspTyrVal 149
DB 835 CTCAGTCAGCGCGTCAACAAACGCTGGGACGCC----- 867
QY 150 ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyThr 169
DB 868 -----GGTATAGTCTGCGTCCGCGCGCAACAGCGCGCGCAACCTACAC 918
QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
DB 919 GTCGCTCACCGCGCGCGGAGCAAGGTCTATACCTCGGTGCA----- 963
QY 190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
DB 964 -----GTTGACAGCAACGACACATCGCCAGCTTCTCCAGCGAGGGA 1005
QY 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229

```


QY 294 -----AlaAspIleGlyLeuGlyTyProAsnGlyAsnGlnGlyTrpGlyArgVa 310
 Db 852 AGATGAATAGCGGATATAGCTACGGTGCA-----GGTAGGT 890
 QY 310 lThrLeuAspIleGlyLeuAsnValAlaTyValAsnGluSerSerLeuSerThrSe 330
 Db 891 TAATGCATACAAAGGTATAAAC-----TACGATAACTATGCAAGCTAGTGTCACTGG 944
 QY 330 rGlnIlyAla-----ThrTySerPheThrAlaThrAlaGlyLysProLe 345
 Db 945 ATATGTTGCCAACAAAGGCGCAACTCACCGATTCGTTATTAGCGAGGTTTCGTCGT 1004
 QY 345 llyIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
 Db 1005 AACTGCCACATTATCTGGGCAATGCCAAT----- 1035
 QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyValGlyAsnAspPh 385
 Db 1036 AGCGACCTTGATCTTACCTCTACGATCCCAATGGAACACGAG---GTGACTACTCTTA 1091
 QY 385 eThrSerProTyAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
 Db 1092 CACCGCTACTAT-----GGATTGGAAGGTTGGTTATTA 1127
 QY 405 nAlaProGlnSerGlyThyTyThrIleGluValGlnAlaTyAsnValProValGlyPr 425
 Db 1128 CAACCCAACTGATGGAACATGACATTAAGGTTGTAAGCTACAGC-----GGAAG 1178
 QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
 Db 1179 TGCMAACTATCAAGTAGATGTGTGAAGT 1206

RESULT 8

US-08-894-818B-4
 ; Sequence 4, Application US/088948:8B
 ; Patent No. 6261822
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAKURA, Hikaru
 ; APPLICANT: MORISHITA, Mio
 ; APPLICANT: YAMAMOTO, Katsuhiko
 ; APPLICANT: MITTA, Masanori
 ; APPLICANT: ASADA, Kiyozo
 ; APPLICANT: TSUNASAWA, Susumu
 ; APPLICANT: KATO, Ikunoshin
 ; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
 ; NUMBER OF SEQUENCES: 42
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Browdy and Neimark
 ; STREET: 419 Seventh Street N.W., Ste. 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: United States of America
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/894,818B
 ; FILING DATE: 20-MAY-1998
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/JP96/03253
 ; FILING DATE: 07-NOV-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 323285/1995
 ; FILING DATE: 12-DEC-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Browdy, Roger L.
 ; REGISTRATION NUMBER: 25,618
 ; REFERENCE/DOCKET NUMBER: TAKAKURA=1

TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 628-5197
 ; TELEFAX: (202) 737-3528
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1566 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: genomic DNA
 ; FEATURE:
 ; OTHER INFORMATION: /note= N at position 1283 is G or T.
 ; US-08-894-818B-4

Alignment Scores: 1.02e-33 Length: 1566
 Pred. No.: 416.50 Matches: 141
 Score: 416.50
 Percent Similarity: 42.55% Conservative: 59
 Best Local Similarity: 30.00% Mismatches: 148
 Query Match: 18.54% Indels: 122
 DB: 3 Gaps: 20

US-09-985-689A-1 (1-434) x US-08-894-818B-4 (1-1566)

QY 12 ValAlaGlnSerSerTyGly-LeuTy-----GlyGlnGlyG1 24
 Db 24 GTCTGCAGCTCAAGTATGGAACCTTACGTTGGAACTTGGGATATGATGGTCTCGAAT 83
 QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMethHisG1 44
 Db 84 CACAATAGGAATAATTGACACTGGAATTCAC-----GCTTCTCATCC 125
 QY 44 uAlaPheArgGlyIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
 Db 126 AGATCTCCAAGGAAGAAGTA-----ATTGGGTGGGTAGATTTCGTCATGG 170
 QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
 Db 171 TAGGAGTTATCCATACGATGACCATGACATGGAACATCATAGCTTCAATAGCAGCTGG 230
 QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
 Db 231 TACTGGAGCAGCAAGTAATGGCAAGTACAGGAATGGCTCCAGGAGCTAAGTGGCGGG 290
 QY 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
 Db 291 AATTAAGTTCTAGGTCCCGATGCTTCTGGAACATATCTACTATAATTAAGGAGCTTGA 350
 QY 112 nThrLeuPheSerGlnAlaTySerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
 Db 351 GTGGCGGTTGTATAACAAAGATAAGTACGGAATTAAGGTCAATTAATCTTCTCTGTTTC 410
 QY 132 a-----AlaValAsnGlyValaTyThrTh 140
 Db 411 AAGCCAGAGCTCAGATGGTACTACGCTCTAAGTCAGGCTGTTAATGCGGCGGATGC 470
 QY 140 rAspSerArgAsnValAspAspTyValArgLysAsnAspMethTrileuPheAlaAl 150
 Db 471 T-----GGATTAGTTGTTGTGTTCCCGC 494
 QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla11 180
 Db 495 TGGAAACAGTGGACCTTACCAAGTATACATCGGTTCTCCAGCAGCTGCAAGCAAGTTAT 554
 QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAs 200
 Db 555 TACAGTTGGAGCC-----GTTGACCAAGTATGA 591
 QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
 Db 582 TGTATTACCAAGCTTCTCAAGCAGAGGCGCAACTGCAGCGCAGGCTTAAGCTGAGGT 641
 QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240

Db 642 TGTGCTCCAGGAACATGATAATTCTCCAGACAGCT-...GGAACTAGCAT 692
QY 240 eTrpAlaAsnHisAspSerLysTyzAlaTyMetGlyGlyThrSerMetAlaThrProIl 260
Db 693 GGGTCAACCAATTAAGACTATTACACAGCAGCTCTGGACATCAATGGCACTCTCA 752
QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLeuThrPr 280
Db 753 CGTAGCTGTATTGACGCCCTCTGTCCAA-...GCACACCC 791
QY 280 olys-...ProSerLeuLeuLysAlaAlaLeuLeuAlaGlyAla-... 293
Db 792 GAGCTGGATCCAGACAAGATAAACAAGCCCTCATAGAACTGCTGATATCGTAAAGCC 851
QY 294 -...AlaAspIleGlyLeuGlyTyzProAsnGlyAsnGlnGlyTyzGlyArgVa 310
Db 852 AGATGAATAGCGCATATAGCTTACCGTGCA-...GGTAGGCT 890
QY 310 lThrLeuAspLysSerLeuAsnValAlaTyValAsnGluSerSerSerLeuSerThrSe 330
Db 891 TAATGATACACAGCTATAAAC-...TACGATAACTATGCAAGCTAGTCTTCACTGG 944
QY 330 rGlnLysAla-...ThrTyzSerPheThrAlaThrAlaGlyLysProle 345
Db 945 ATATGTTGCCAACAAAGGAGCCAAACTCACAGTTCGTTATTAGCGAGCTTCGTTCT 1004
QY 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1005 AACTGCCACATATATCTGGACAACTGCAAT-... 1035
QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyzValGlyAsnAspPh 385
Db 1036 -AGCGACTTGATCTTACCTACGATCCATGGAACACAG-...GTTGACTACTCTTA 1091
QY 385 eThrSerProTyzAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1092 CACCGCTACTAT-...GGATTCGAAAAGGTTGGTTATTA 1127
QY 405 nAlaProGlnSerGlyThrTyzThrIleGluValGlnAlaTyzAsnValProValGlyPr 425
Db 1128 CAACCCAACTGATGGAAACATGGAACATTAAGTTGTAAGCTACAGC-...GGAG 1178
QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1179 TGCAAACTATCAAGTAGATGTGTAAGT 1206

RESULT 9

US-08-894-818B-34
; Sequence 34, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1962 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-894-818B-34
Alignment Scores:
Pred. No.: 1,46e-33 Length: 1962
Score: 416.50 Matches: 141
Percent Similarity: 42.55% Conservatve: 59
Best Local Similarity: 30.00% Mismatches: 148
Query Match: 18.54% Indels: 122
DB: 3 Gaps: 20
US-09-985-689A-1 (1-434) x US-08-894-818B-34 (1-1962)
QY 12 ValAlaGlnSerSerTyzGly-LeuTy-...-GlyGlnGlyGl 24
Db 420 GTCTGCAGCTCAAGTTATGCACTTACGTTTGGAACTGGGATATGATGGTTCTGGAAT 479
QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisGl 44
Db 480 CACAATAGGAATAATTGACACTGGAATTGAC-...-GCTTCTCATCC 521
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsn- 63
Db 522 AGATCTCCAGGAAAGCTA-...-ATTGGGTGGGTAGATTGTCATG 566
QY 64 -...-AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGl 78
Db 567 TAGGATATCCATACCATGACCATGGAACTGCACTGATGCTTCAATAGCAGCTGG 626
QY 78 yAsnGlySerThrAsn-...LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGCAAGTACAAGGAATGGCTCCAGGAGCTAAGCTGGCGGG 686
QY 94 eGlnSerIleMet-...-AspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGl 112
Db 687 AATTAAGGTTCTAGTGCCGCTGTTCTGGAAGCATATCTACTATAATTAAGGAGTTGA 746
QY 112 nThrLeuPheSerGlnAlaTyzSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 747 GTGGCCCGTTGATAACAAGATAAGTACGGAATTAAGCTATTAATCTTCTCTGTTTC 806
QY 132 a-...-AlaValAsnGlyAlaTyzThrTh 140
Db 807 AAGCCAGAGCTCAGATGCTAGCTGCTTAAGTCAAGCTGTTAATGAGCTGGGATGC 866
QY 140 rAspSerArgAsnValAspAspTyzValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 867 T-...-GGATTAGTTGTTGGTTGCCGC 890
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIl 180

Db 891 TGGAAACAGTGGACCTAACAACTATACAACTCGTTCTCCAGCAGCTGCAACCAAGTTAT 950
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerThrAlaAspAsnIleAs 200
Db 951 TACAGTTGGAGCC-----GTGCAACAGTATGA 977
Qy 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 978 TGTATTAACAAGCTTCTCAAGCAGAGGCGCACTGACAGCGCAGGCTTAAGCTGAGGT 1037
Qy 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 1038 TGTCTCTCCAGAACTGGATAATGCTGCCAGACAACT-----GGAACCTAGCAT 1088
Qy 240 eThrAlaAsnHisAspSerLysThrAlaThrMetGlyGlyThrSerMetAlaThrProIl 260
Db 1089 GGGTCAACCAATTAATGACTATTACACAGCAGCTCTGGACATCAATGGCACTCTCA 1148
Qy 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 1149 CGTAGCTGTATGTCAGCCCTCTTCTCCAA-----GCACACCC 1187
Qy 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 1188 GAGCTGGACTCAGACAAAGTAAACAGCCCTCTATAGAACTGCTGATATCGTAAAGCC 1247
Qy 294 -----AlaAspIleGlyLeuGlyTyPrProAsnGlyAsnGlnGlyTyPrGlyArgVa 310
Db 1248 AGATGAATAGGCGATATAGCTCGGTGCA-----GGTAGGGT 1286
Qy 310 lThrLeuAspLysSerLeuAsnValAlaThrValAsnGluSerSerSerLeuSerThrSe 330
Db 1287 TAATGCATCAAGGCTATAAAC-----TACGATTAACATGCAAGAGTAGTGTTCACCTGG 1340
Qy 330 rGlnLysAla-----ThrTySerPheThrAlaThrAlaGlyLysProLe 345
Db 1341 ATATGTTGCCAAAGAGGCGACCAACTCACCAGTTCGTTATTAGCGGAGCTTCGTTCTGT 1400
Qy 345 uLysIleSerLeuValThrSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1401 AACTGCCACATTATATCTGGGCAATGCCAAT----- 1431
Qy 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyValGlyAsnAspPh 385
Db 1432 -AGCGACCTTGATCTTACCTCTACGATCCCAATGGAACCCAG--GTTGACTACTCTTA 1487
Qy 385 eThrSerProTyAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1488 CACCGCCTACTAT-----GGATTGCAAAAGGTTGGTTATTA 1523
Qy 405 nAlaProGlnSerGlyThrTyThrIleGluValGlnAlaThrAsnValProValGlyPr 425
Db 1524 CAACCCACATGATGGAACTGACATGACATTAAGTTGTAGCTACAGC-----GGAAG 1574
Qy 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1575 TGCAAACTATCAAGTAGATGTGGTAGT 1602

RESULT 10

US-09-445-472-15
; Sequence 15, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikumoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06

; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-15
Alignment Scores:
Pred. No.: 1,46e-33 Length: 1962
Score: 416.50 Matches: 141
Percent Similarity: 42.55% Conservative: 59
Best Local Similarity: 30.00% Mismatches: 148
Query Match: 18.54% Indels: 122
DB: 4 Gaps: 20

US-09-985-689A-1 (1-434) x US-09-445-472-15 (1-1962)

Qy 12 ValAlaGlnSerSerTyrgly-LeuTyrr-----GlyGlnGlyG 24
Db 420 GTCTGCAGCTCAAGTTATGCAACTTACGTTTGGAACTTGGGATATGATGTTCTGGAAT 479
Qy 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMethIleGl 44
Db 480 CCAATAGGAATAATTGACACTGGAATGAC-----GCITCTCATCC 521
Qy 44 uAlaPheArgGlyLysIleThrAlaLeuTyrrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAAGAAAGTA-----ATTGGTGGGTAGATTGTTCAATGG 566
Qy 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGl 78
Db 567 TAGGAGTTATCCATACGATGACCATGACATGGAACATCATGAGTCTCAATAGCAGCTGG 626
Qy 78 YAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGGCAAGTACAGGNAATGGTCCAGGAGCTTAAGCTGCGGG 686
Qy 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyLeuGlyLeuProSerAsnLeuGl 112
Db 687 AATTAAGTTCTAGTGCCGATGGTCTCTGGAACATATCTACTATAATTAAAGGGAGTTGA 746
Qy 112 nThrLeuPheSerGlnAlaTyrrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 747 GTGGGCGGTTGATAACAAAGATAGTACGGAATTAAGTCATTAACTCTTCTCTGTTTC 806
Qy 132 a-----AlaValAsnGlyAlaTyrrThrTh 140
Db 807 AAGCCAGAGCTCAGATGGTACTGACGCTCTTAAGTCAGGCTGTTTAATGACGCGTGGGATGC 866
Qy 140 rAspSerArgAsnValAspAspTyrrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 867 T-----GGATTAGTTGTTGTTGTTCCCGC 890
Qy 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIl 180
Db 891 TGGAAACAGTGGACCTTAACAAGTATACAATCGGTTCTCCAGCAGCTGCAAGCAAGTTAT 950
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrrAlaAspAsnIleAs 200
Db 951 TACAGTTGGAGCC-----GTTGACAAGTATGA 977
Qy 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 978 TGTATTAACAAGCTTCTCAAGCAGAGGCGCACTGACAGCGCAGGCTTAAGCTGAGGT 1037
Qy 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 1038 TGTGTCTCCAGAACTGGATAATGCTGCCAGACAACT-----GGAACCTAGCAT 1088

240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrPro11 260
 1089 GGTACCAACAAATTAAGTATATACACAGCAGCTCTGGACATCAATGGCACTCTCA 1148
 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
 1149 CGTAGCTGTATTGAGCCCTCTTGCTCCA
 280 olys-----ProSerLeuLeuLysAlaAlaLeuLysAlaGlyAla----- 293
 1188 GAGCTGGACTCAGACAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 1247
 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlyGlyTyrGlyArgVa 310
 1248 AGATGAATAGCGGATATAGCTACGGTCA-----GGTAGGGT 1286
 310 ThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330
 1287 TAATGATCAACAGGCTATAAAC-----TACGATAACTATGCAAGAGCTAGTGTTCAC 1340
 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
 1341 ATATGTTGCCAACAAAGGCGCAAACTCACCAGTTGCTTATTAGCGGAGCTTCGTCGT 1400
 345 uLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
 1401 AACTGGCCACATTAATCTGGGCAATCCCAAT----- 1431
 365 LAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
 1432 -ACGGACCTTGATCTTACCTACATCCCAATGGAACAGC---GTTGACTACTCTTA 1487
 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
 1488 CACGGCTACTAT-----GGATTGAAAGAGGTGTTATTATTA 1523
 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
 1524 CACCCCACTGATGGACATGGACATTAAGTTGTAAGCTACAGC-----GGAAG 1574
 425 oGlnThrPheSerLeuAlaIleValAsn 434
 1575 TGCAAACTATCAAGTAGATGTGGTAAGT 1602

RESULT 11

US-08-894-818B-6
 ; Sequence 6, Application US/08894818B
 ; Patent No. 6261822
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAKURA, Hikaru
 ; APPLICANT: MORISHITA, Mio
 ; APPLICANT: YAMAMOTO, Katsuhiko
 ; APPLICANT: MITTA, Masenori
 ; APPLICANT: ASADA, Kiyozo
 ; APPLICANT: TSUNASAWA, Susumu
 ; APPLICANT: KATO, Ikunoshin
 ; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
 ; NUMBER OF SEQUENCES: 42
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Browdy and Neimark
 ; STREET: 419 Seventh Street N.W., Ste. 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: United States of America
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/894,818B

; FILING DATE: 20-MAY-1998
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/JP96/03253
 ; FILING DATE: 07-NOV-1996
 ; PRIOR APPLICATION DATA: JP 323285/1995
 ; APPLICATION NUMBER: 12-DEC-1995
 ; FILING DATE: 12-DEC-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Browdy, Roger L.
 ; REGISTRATION NUMBER: 25,618
 ; REFERENCE/DOCKET NUMBER: TAKAKURA=1
 ; TELEPHONE: (202) 628-5197
 ; TELEFAX: (202) 737-3528
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1977 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: genomic DNA
 ; US-08-894-818B-6

Alignment Scores:

Pred. No.: 3,36e-32 Length: 1977
 Score: 403.50 Matches: 134
 Percent Similarity: 43.07% Conservatives: 68
 Best Local Similarity: 28.57% Mismatches: 147
 Query Match: 17.96% Indels: 120
 DB: 3 Gaps: 20

US-09-985-689A-1 (1-434) x US-08-894-818B-6 (1-1977)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGly 24
 Db 420 GTCTGCAGCTCAAGTTATGCACTTACGTTGGAACTGGATATGATGTTCTGGAAT 479
 QY 24 nileValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG 44
 Db 480 CACATAGGAATATTGACACTGGAATTGAC-----GTTCTCATCC 521
 QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
 Db 522 AGATCTCCAGGAAAGTA-----ATTGGGTGGGTAGATTTTGTCAATGG 566
 QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeu 78
 Db 567 TAGGAGTTATCCATAGTACGATGACATGGACATGGAACTCATGTAGCTTCAATAGCAGCTGG 626
 QY 78 YAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
 Db 627 TACTGGAGCAGCAAGTAATGGCAAGTACAGGAATGGCTCCAGGAGCTAAGCTGGCGGG 686
 QY 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeu 112
 Db 687 AATTAAGTTCTTAGTGGCCGATGTTCTTGGAACTATCTACTATAATTAAGGAGTGA 746
 QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIle----- 125
 Db 747 GTGGGCCGTTGATAACAAAGATAAGTACGGAATTAAGGTCAATCTTCTCTGGTTC 806
 QY 126 -----HisThrAsnSerTrpGlyAlaValAlaValAsnGlyAlaTyrThr 140
 Db 807 AAGCCAGAGCTCCGACGGAACCGACTCCCTCAGTCAGCCGCTCACACGCTCGGAGCC 866
 QY 140 rAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
 Db 867 C-----GGTATAGTAGTCTCGCTCGCGCC 890
 QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI 180
 Db 891 CGGCAACACGGCGCGGAACACCTACACCGTCCGCTCACCCCGCCGCGGAGCAAGGTCTAT 950

```
QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 951 AACCGTCGGTGA-----GTTGACGACGACGA 977
QY 200 nHisValAlaGlnPheSerArgGlyProThrIleAspGlyArgIleIysProAspVa 220
Db 978 CAACATCCCGAGCTTCTCCAGCAGGGACCGACCGGCGGAGGCTCAAGCGGAAGT 1037
QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPh 240
Db 1038 GTCCGCCCGCGGTGACATCATAGCCCGGCGCCAGC-----GGAACACGAT 1088
QY 240 eTrpAlaAsnHisAspSerIleTyrAlaTyrMetGlyThrSerMetAlaThrProIl 260
Db 1089 GGGCACCCCGGATAAACGACTACTACCAAGCGCTCTGGAACGAGCATGGCCACCGCGCA 1148
QY 260 eValAlaGly---AsnValAlaGlnLeuArgGluHisPheValIysAsnArgIleTh 279
Db 1149 GTTTCGGCGGTGGCGGCTCATCTCCAGGCCAC-----1185
QY 279 rProLys-----ProSerLeuLeuIleAlaAlaLeuIleAlaGlyAla----- 293
Db 1186 -CCGAGCTGGACCCCGGACCAAGGTGAGACCGCCCTCATCGAGACCGCGACATAGTCG 1244
QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyAr 309
Db 1245 CCCCAGGAGATAGCGGACATCGCTACGGTGG-----GGTAG 1283
QY 309 gValThrLeuAspIysSerLeu-----AsnValAlaTyrValAsnGluSerSe 326
Db 1284 GGTGAACGTCACAGGCGCATCAAGTACGACGACTACGCCAAGCTCACCTTCACGGGCTC 1343
QY 326 rLeuSerThrSerGlnIysAlaThrTyrSerPheThrAlaThrAlaGlyIysProLeuLy 346
Db 1344 CCGCCCGCAGGAAGGAGCGCCACACCTTCGACGCGCGCCACCTTCGTGAC 1403
QY 346 sIleSerLeuValTyrSerAspAlaProAlaSerThrThrAlaSerValThrLeuValAs 366
Db 1404 CGCCACCTCTACTGGAC-----ACGGCTCGAG 1433
QY 366 nAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPheTh 386
Db 1434 CGACATCGACCTCTACCTCTACGACCCCAACGGAACGAG---GTTGACTACTCTCTACAC 1490
QY 386 rSerProTyrAsnAspAsnTyrAspGlyArgAsnAsnValGluAsnValPheIleAsnAl 406
Db 1491 CCCCTACTAC-----GGCTTCGAGAGGTCGGCTACTACAA 1526
QY 406 aProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyProGl 426
Db 1527 CCGGACCGCGGAACTCGACGCTCAAGTCTGCTAGCTACAAG-----GGCGGCGC 1577
QY 426 nThrPheSerLeuAlaIleValAsn 434
Db 1578 GAACCTACAGGTCGACGTCGTGAGC 1602

RESULT 12
US-09-000-016-3
; Sequence 3, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
```

```
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000,016
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:
```

```
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2539 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
US-09-000-016-3
```

```
Alignment Scores:
Pred. No.: 1,51e-26 Length: 2539
Score: 351.00 Matches: 131
Percent Similarity: 42.63% Conservative: 54
Best Local Similarity: 30.18% Mismatches: 159
Query Match: 15.62% Indels: 90
DB: 3 Gaps: 16
```

US-09-985-689A-1 (1-434) x US-09-000-016-3 (1-2539)

```
QY 2 AspValAlaArgGlyIleValIysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
Db 953 GACACGTCTCGTGGGAGATCGGCGCCCAAGGCGTGTCTCGCC---GGCTACGACGCG 1009
QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 41
Db 1010 AAGGCGTGAAGATCGCGTCTCTGGACACCGGTGTGACACGAGC-----1054
QY 42 MetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGGACCTGAAGGCGCGGTGACCGGTCCAGAACTTACCCCGCGCGCCGCGC 1111
QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
Db 1112 GCCGGGACAGAGGTGGGCGCACCCACGCTCGCTCGATCGGCGCGGCGCACGCGCGCC 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CAGTCCAAGGGCAAGTACAGGCGCTCGACCCCGCGCGCGATCTCTCAACGCGAAGGTC 1231
QY 98 MetAspSerGlyGlyLeuGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db 1232 CTGACGACTCCGCT-----TTGCGCGACGACTCCGGCATCTCGCGCGCATGGAGTGG 1285
QY 118 AlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAla 137
```

Db 1286 GCGCCGCGGAGGCGCGCGCTCGTCACCATGAGCTGGCC-----GGCATG 1333
Qy 138 TyrThrAspSerArgAsnValAspAspTyrValArgLys-----AsnAspMetThr 155
Db 1334 GACACCGGAGACCGACCGCTGGAGCGCGCTGCACAAGCTGCTCGCGCGAAGAGGC 1393
Qy 156 IleLeuPhe-----AlaAlaGlyAsnGluGlyProAsnGlyThrIleSerAlaPro 173
Db 1394 GTCCTGTTCGCCAFCGCGGCGCGGCAACGAGGCGCGGAG-----TGATCGGTTCGCC 1447
Qy 174 GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGly 193
Db 1448 GGCAGCGCGGAGCGCGCTCACCGCTCGCGCC----- 1480
Qy 194 SerTyrAlaLeuAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro-----ThrLys 212
Db 1481 -----GTCGACGACCAAGCAAGCTCGCGACTTCCTCCACCGCGCGCGCTCGGC 1534
Qy 213 AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSer 232
Db 1535 GACGCGCATCAAGCGGAGCTCACCGCTCCCGGCTGGACATCAGCGCGCTCGGC 1594
Qy 233 SerLeuAlaProAspSerSerPheTyrAlaAsnHisAspSerLysTyrAlaTyrMetGly 252
Db 1595 GAGGCGCAACGACATCGCGGAGGCTCGGTGAGGACCGCGCGCTACATGACCATCTCC 1654
Qy 253 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe 272
Db 1655 GGCAGCTGATGGAGCGCGCGACGTCGCGGCGCGCGCGCTCCGTAAGAGCAG--- 1711
Qy 273 ValLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu----- 289
Db 1712 -----CACCCGAGCTGGACCTCCCGCGAAGTGAAGGCGCG 1747
Qy 290 IleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArg 309
Db 1748 CTCACCGCTCCACCAAG---GGCGCAAGTACACCCGCTCGAGCAGGCTTCGGGCGCG 1804
Qy 310 ValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerLeuSer--- 328
Db 1805 ATCCAGCGCGACAGCGCTCCAGCAGACCGTATCCGCGCGCGCTCGGTGAGCTTC 1864
Qy 329 -----ThrSerGlnLysAlaThrTyrSer 336
Db 1865 GCGCTCCAGCAGTGGCGCGCACACGACGAGCGCGTCCACCAAGCAGCTGACCTACCGC 1924
Qy 337 PheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTyrPheAsp----- 353
Db 1925 AACCTCGGACCCAGGAGCTCACGCTGAGCTGAGCTGACCGCGCGCGCGCGCGCG 1984
Qy 354 -----AlaProLysSer 357
Db 1985 AAGCGCGCGCGCGCGCTTTCACGCTGGCGCGCACACCGGTGACCGTCCGCGCGCG 2044
Qy 358 ThrThrAlaSerValThrLeuValAsnAspLeuAspLeu----- 370
Db 2045 GGCAGCGCTCCGTCGACATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2104
Qy 371 -----ValIleThrAlaProAsnGlyThrGlnTyrVal 381
Db 2105 TACTCGGCTACGTGGTGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2146

RESULT 13

US-09-514-340-3
; Sequence 3, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: AKIRA ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPIRIDINE DERIVATIVES
; ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/514,340
FILING DATE: 28-Feb-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/000,016
FILING DATE: January 30, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2539 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-514-340-3

Alignment Scores:
Pred. No.: 1,51e-26 Length: 2539
Score: 351.00 Matches: 131
Percent Similarity: 42.63% Conservative: 54
Best Local Similarity: 30.18% Mismatches: 159
Query Match: 15.62% Indels: 90
DB: 4 Gaps: 16

US-09-985-689A-1 (1-434) x US-09-514-340-3 (1-2539)

Qy 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
Db 953 GACAGCTCGTCGCGCAGATCGGCGCCCAAGGCGTGTCCGCC---GGCTACGACGCG 1009
Qy 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 41
Db 1010 AAGGCGGTGAAGATCGCGCTCTCGACACCGGTGTGACACGAGC----- 1054
Qy 42 MetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCGGACCTGAAGGCGCGGTGACCGGTCCAGAACTTCACCGCGCGCGCGC 1111
Qy 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
Db 1112 GCGCGCAGAGTGGCGCAGCGCACACCGCTCGCTCGATCGCGCGCGCGCGCGCGCC 1171
Qy 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CAGTCCAAAGGCAAGTACAGGCGCGTGCACCGCGCGCGCGGTCTCTCAACGCGAGGTC 1231

QY 98 MetAspSerGlyGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db 1232 CTCGACGACTCGGT-----TTGGCGAGCACTCGGCATCTCGCGGATGGAGTGG 1285
QY 118 AlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAla 137
Db 1286 GCGGCGCGCGAGCGCGCGACCTCGTCACCATGAGCTGGGC-----GGCATG 1333
QY 138 TyrThrThrAspSerArgAsnValAspTyrValArgLys-----AsnAspMetThr 155
Db 1334 GACACACCGAGACCGACCGCTGGAGCGGGCTCGACAGCTCTCGCGGAGAGGCG 1393
QY 156 IleLeuPhe-----AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaPro 173
Db 1394 GTCTGTTCGCCATCGCGCGCGGCAACGAGGGCCCGGAG-----TCGATCGGTTCGCC 1447
QY 174 GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGly 193
Db 1448 GGCAGCGCGGACCGCGCGCTCACCTCGCGCGC----- 1480
QY 194 SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLys 212
Db 1481 -----GTCACACAGCAAGCAAGCTCGCGGACTTCTCTCCACCGCGCGCGCTCGGC 1534
QY 213 AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSer 232
Db 1535 GACGGCGCATCAAGCGCGGACGTCACGCTCCGCGGTGGACATCAGCGCGCGCTCGCGG 1594
QY 233 SerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGly 252
Db 1595 GAGGGCAACGACATCGCGCAGGAGTGGTGGAGGACCGCGCGGTACATGACCATCTCC 1654
QY 253 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe 272
Db 1655 GGCAGCTCGATCGGACCGCGCGACGTCGCGCGCGCGCGCGCTCTCTGAAGCAGCAG--- 1711
QY 273 ValLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu----- 289
Db 1712 -----CACCCGAGTGGACCTCCGCGGACTGAGGCGCG 1747
QY 290 IleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArg 309
Db 1748 CTCACCGGTCCACCAAG---GGCGCAAGTACACCCCGTTCGAGCAGGGTTCGGGCGCG 1804
QY 310 ValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerLeuSer--- 328
Db 1805 ATCAGCGCGACAAAGGGCTCCAGCAGCGTATCGCGACCGGTCTCGGTGACCTTC 1864
QY 329 -----ThrSerGlnLysAlaThrTyrSer 336
Db 1865 GCGGTCCAGCAGTGGCGGCACACCGACGCGCGGTCTACCAAGCAGCTGACCTACCGC 1924
QY 337 PheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp----- 353
Db 1925 AACCTCGGACCCAGGAGCTGACGCTGAGCTGACGCGCGACCGCGCCACCGCCAGGCG 1984
QY 354 -----AlaProAlaSer 357
Db 1985 AAGCGCGCCCGCGGCGCTTCTTCACGCTGGCGCGCACACCGGTGACCGTCCCGGCGGC 2044
QY 358 ThrThrAlaSerValThrLeuValAsnAspLeuAspLeu----- 370
Db 2045 GGCAGCGCGCTCCGTCGACATGACCGCGCGACACCGCGCTCGCGCGCAGCGGTGAGCGCG 2104
QY 371 -----ValIleThrAlaProAsnGlyThrGlnTyrVal 381
Db 2105 TACTCGCGCGTACGTGGTCCGACCGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGG 2146

RESULT 14

US-09-000-016-1

; Sequence 1, Application US/09000016

; Patent No. 6143541

; GENERAL INFORMATION:

APPLICANT: Akira ARISAWA et al.
TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROXYRIDINE DERI
TITLE OF INVENTION: ITS EXPRESSION PRODUCT
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wengeroth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000,016
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2809 base pairs
TYPE: nucleic acid
STRAINEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
ORIGINAL SOURCE:
ORGANISM: Streptomyces antibioticus
STRAIN:
FEATURE:
NAME/KEY: CDS
LOCATION: 338..2539
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: CDS
LOCATION: 2540...2809
IDENTIFICATION METHOD: P
US-09-000-016-1

Alignment Scores:
Pred. No.: 5,9e-26 Length: 2809
Score: 346.00 Matches: 129
Percent Similarity: 42.00% Conservative: 52
Best Local Similarity: 29.93% Mismatches: 167
Query Match: 15.40% Indels: 84
DB: 3 Gaps: 13

US-09-985-689A-1 (1-434) x US-09-000-016-1 (1-2809)

QY 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
Db 953 GACAGCTCCGTCGGCGAGATCGCGCGCCCAAGCGGTGTCGCGC---GGTACGACGCG 1009
QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSer 41
Db 1010 AAGCGCGTGAAGATCGCGCGTCTCGGACCGCGTGTGACACGAGC----- 1054

Qy 42 MetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGACCTGAAGGCGCGGTGACCGCGTCCAGAACTTCAACCGCGCGCCGCGC 1111
Qy 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
Db 1112 GCCGGGACAAGGTGGGCGCACGCCACCGTCGCTCGATCGCGCGCGGCGACGGCGCC 1171
Qy 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CATCTCAAGGGCAAGTACAAGGCGGTGCGACCGCGCGCGCGATCTCAACGGCAAGGTC 1231
Qy 98 MetAspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db 1232 CTCGACGACTCCGGT-----TTCCGGCGACGACTCCGGCATCTCCCGGCATGGAGTGG 1285
Qy 118 AlaTySerAlaGlyAlaAlaGlyIle-HisThrAsnSerTrpGlyAlaAlaValAsnGlyAl 137
Db 1286 GCGGCGCGCGAGGCGCGGCGTGTCAATGATGCTGGCGGCGCATGGACACACCGCGAG 1345
Qy 137 aTyThrThrAspSerArgAsnValAspAspTyValArgLysAsnAspMetThrIleLe 157
Db 1346 ACCGACCGGTGGAGGCGGGG-GTCGACAAGCTGTCCGCGCGAGAGGGCGTCTGTTCGC 1404
Qy 157 uPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaTy 177
Db 1405 CATCGCGCGCGCAACAGAGGCGCGGAG-----TCGATCGGTTCGCGCGGCGCGGA 1458
Qy 177 sAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAs 197
Db 1459 GCGCGCCCTCACCGTCGCGGCC-----GTCCGA 1485
Qy 197 pAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLysAspGlyArgI 216
Db 1486 CGACAAGGACAAGTCTCCGACTTCTCTCCACCGGCGCCGCTCGGCGAGCGGCCAT 1545
Qy 216 eLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPr 236
Db 1546 CAAGCGGAGGTCAACCGCTCCGCGGTGGACATCACGCGCGCTCGGCGAGGGGCAACGA 1605
Qy 236 oAspSerSerPheTrpAlaAsnHisAspSerLysTyAlaTyMetGlyGlyThrSerMe 256
Db 1606 CATCGCGCAGAGGTCCGTGAGGAGCGCGCGCTACATGACCATCTCCGCGCAGTCGAT 1665
Qy 256 tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAr 276
Db 1666 GCGACCCCGCACGTCCGCGGCGCGGCGCCCTCTGAAGCAGCAG-----1711
Qy 276 gGlyIleThrProLysProSerLeuLysAlaAlaLeu-----IleAlaGlyAl 293
Db 1712 -----CACCCCGACTGGACCTCCGCGGAACTGAAGGCGCGCTCACCGGCTC 1758
Qy 293 aAlaAspIleGlyLeuGlyTyProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAs 313
Db 1759 CACCAAG---GGCGGCAAGTACACCCGTTTCGAGCAGGGTTCGGGCGCGATCCAGGCGCA 1815
Qy 313 pLysSerLeuAsnValAlaTyValAsnGluSerSerSerLeuSer-----328
Db 1816 CAAGCGCTCAGCAGACCGTATCGCGGACCGCGCTCTCGGTGAGCTCGGCGGTCCAGCA 1875
Qy 329 -----ThrSerGlnLysAlaThrTySerPheThrAlaTh 340
Db 1876 GTGGCGGCACACGACGAGCGGTCCACCAAGCAGCTGACCTACCGCAACCTCGGCAC 1935
Qy 340 rAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp-----353
Db 1936 CCAGGAGTCACGCTGAAGTGAAGTCGACGCGCGCACCGCAACCCCAAGGCGAAGCGGCCCC 1995
Qy 354 -----AlaProAlaSerThrThrAlaSe 361
Db 1996 GCGGGGCTTCTTACGCTGGGCGCCACCGGTGACCGTCCGCGGGCGGCGAGCGCCCTC 2055
Qy 361 rValThrLeuValAsnAspLeuAspLeu-----Va 371

RESULT 15

US-09-514-340-1
; Sequence 1, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE
; IIS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/514,340
; FILING DATE: 28-Feb-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/000,016
; FILING DATE: January 30, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng
; REGISTRATION NUMBER: 40,949
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2809 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces antibioticus
; STRAIN: <Unknown>
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 338..2539
; IDENTIFICATION METHOD: E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2540...2809
; IDENTIFICATION METHOD: P
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-514-340-1

Alignment Scores:
Pred. No.: 5.9e-26 Length: 2809
Score: 346.00 Matches: 129
Percent Similarity: 42.00% Conservative: 52
Best Local Similarity: 29.93% Mismatches: 167
Query Match: 15.40% Indels: 84
DB: 4 Gaps: 13

US-09-985-689A-1 (1-434) x US-09-514-340-1 (1-2809)

```
Qy 2 AspValAlaArgGlyLeuValAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
Db 953 GACACGTCGTCGGGAGATCGCGCCCAAGGGGTGTGTCGCC---GGTACGACGGC 1009
Qy 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSer 41
Db 1010 AAGGGGTGAAGATCGCGTCTGGACACCGGTGTGACACGAGC----- 1054
Qy 42 MetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGGAGCTCAAGGGCGGGGTGACCGGTCCAGAACTTCACCGCGCGCGCCGCG 1111
Qy 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
Db 1112 GCGGGGACAAAGTGGGCGACCGCACCCACGTCGCTCGATCGCGCGGGGCACGGGCGCC 1171
Qy 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CAGTCCAAAGGGCAAGTACAAAGGGCGTGCACCGCGCGCGCATCTCAACGGCAAGGTC 1231
Qy 98 MetAspSerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db 1232 CTCGACACTCGGT-----TTCGGGACGACTCCGGCATCTCCCGCGCATGGAGTGG 1285
Qy 118 AlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAlaValAsnGlyAl 137
Db 1286 GCGCGCGCGAGCGCGCGACGTCGTCAACATGAGCTGGGCGGCATGACACACCGGAG 1345
Qy 137 aTyThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLe 157
Db 1346 ACGGCGCGCTGGAGCGCGCG-GTCGACAAAGTGTCCGCGCGAGAGGGCGCTCTTCGC 1404
Qy 157 uPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaL 177
Db 1405 CATCGCGCGCGCAACAGAGGGCGCGGAG-----TCGATCGGTTCGCGCGCGCGCA 1458
Qy 177 ssnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAs 197
Db 1459 CGCGCGCGCTCACCGTCGCGCGC-----GTGCA 1485
Qy 197 pAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLysAspGlyArgI 216
Db 1486 CGACAGGACAGCTCGCGCATCTCTCTCCACCGCGCCCGCCCTCGCGCGCGCGCAT 1545
Qy 216 eLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPr 236
Db 1546 CAAGCGGACGTCACCGCTCCCGGGGTGGACATCACGCGCGCTCGCGGGGCGCAACGA 1605
Qy 236 oAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMe 256
Db 1606 CATCGCGCGAGGTGCGTGAGGGACCGCGCGGTACATGACCATCTCCGGCACGTGAT 1665
Qy 256 tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAr 276
Db 1666 GCGGACCGCGCGAGTCCGCGGGCGCGCGCTCTGAAAGCAGCAG----- 1711
Qy 276 gGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu-----IleAlaGlyAl 293
Db 1712 -----CACCCGACTGGACCTCCCGCACTGAAGGGCGCGCTCACCGGCTC 1758
Qy 293 aAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAs 313
Db 1759 CACCAAG---GGCGGCAAGTACACCCCGTTCGAGCAGGGTTCGGGCGCGATCCAGCGCGA 1815
Qy 313 pLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSer----- 328
Db 1816 CAAGGGCGCTCCAGCAGACCGGTATCGCGACCCCGGTCTCGGTGAGCTTCGGCGTCCAGCA 1875
Qy 329 -----ThrSerGlnLysAlaThrTyrSerPheThrAlaTh 340
Db 1876 GTGGCGCGCACCGACGAGCGGTGTACCAAGCAGCTGACTTACCGCAACCTCGGCAC 1935
```

```
Qy 340 rAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp----- 353
Db 1936 CCAGGACGTCACTGAGCTGACGTGACCGCCACCGACCCCAAGGGCAAGCGCGCCCC 1995
Qy 354 -----AlaProAlaSerThrThrAlaLe 361
Db 1996 GCGGGGTCTTTCACGCTGGCGCCACCACCGGTGACCGTCCCGCGCGGGCGGCGCGCTC 2055
Qy 361 rValThrLeuValAsnAspLeuAspLeu-----Va 371
Db 2056 CGTCGACATGACCGCGGACACCGCGGTCTGGCGGCGACCGGTGAGCGGCGGTACTTCGGCGTA 2115
Qy 371 lIleThrAlaProAsnGlyThrGlnTyrVal 381
Db 2116 CTGGTCTGCCACCGGGCGGGGCGAGCGGTC 2146
```

Search completed: March 16, 2004, 01:18:00
Job time : 118 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 16, 2004, 01:15:59 ; Search time 359 Seconds
(without alignments)
4450.734 Million cell updates/sec

Title: US-09-985-689a-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAINVPVGQTFSLAIVN 434

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09985689/runat_10032004_112305_19840/app_query.fasta_1.583
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=epct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODELOCAL -OUTFMT=pcio -NOR=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USERS=US09985689.ecgn_1_164_brunat_10032004_112305_19840
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARG TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09D_PUBCOMB.seq:
13: /cgn2_6/ptodata/1/pubpna/US09E_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

ALIGNMENTS

RESULT 1

US-10-385-662-1
; Sequence 1, Application US/10385662
; Publication No. US20040002432A1
; GENERAL INFORMATION:
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: SAITO, KAZUHIRO
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: IZAWA, YOSHIFUMI
; APPLICANT: SAKKI, KATSUHIISA
; APPLICANT: KOBAYASHI, TOHRU
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: Alkaline protease
; FILE REFERENCE: 234938USO
; CURRENT APPLICATION NUMBER: US/10385,662
; CURRENT FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: JP 2002-081428
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: JP 2002-165987
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: JP 2002-304230
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: JP 2002-304231

Sequence 1, Appli
Sequence 11, Appli
Sequence 2, Appli
Sequence 15, Appli
Sequence 5701, Ap
Sequence 3306, Ap
Sequence 1, Appli
Sequence 1, Appli
Sequence 1934, Ap
Sequence 5, Appli
Sequence 113, Appl
Sequence 2, Appli
Sequence 29, Appli
Sequence 33, Appli
Sequence 1, Appli
Sequence 5, Appli
Sequence 5384, Ap
Sequence 112429,
Sequence 1938, Ap
Sequence 1935, Ap
Sequence 1, Appli
Sequence 17, Appli
Sequence 17, Appli
Sequence 34, Appli
Sequence 11, Appli
Sequence 12, Appli
Sequence 6, Appli
Sequence 20, Appli
Sequence 20, Appli
Sequence 6, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 6, Appli
Sequence 7, Appli
Sequence 1, Appli
Sequence 30, Appli
Sequence 5689, Ap
Sequence 9, Appli
Sequence 13, Appli

; PRIOR FILING DATE: 2002-10-18
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 1305
 ; TYPE: DNA
 ; ORGANISM: Bacillus sp. KSM-KP43
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1305)
 ; OTHER INFORMATION:
 ; US-10-385-662-1
 Alignment Scores:
 Pred. No.: 8,41e-242 Length: 1305
 Score: 2247.00 Matches: 434
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 15 Gaps: 0
 US-09-985-689A-1 (1-434) x US-10-385-662-1 (1-1305)
 QY 1 AsnAspValalaArgGlyLeuValaAspValaAlaGlnSerSerTyGlyLeuTyr 20
 DB 1 AATGATGTTGCGCGTGAATTTCAAGCGGATGGCTCAGAGCAGCTACGGGTGTAT 60
 QY 21 GlyGlnGlyGlnIleValaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 DB 61 GGCAAGGACAGATCGTAGCGTTCGCCGATACAGGCTTGATACAGTTCGCAATGACAGT 120
 QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyAlaLeuGlyArgThrAsn 60
 DB 121 TCGATCATGAAGCCCTTCGCGCGGAAATTTCTGCATTATATGCAATTCGGACGGAGAT 180
 QY 61 AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 DB 181 ATGGCAATGATACGATGGTTCATGTCATGTCGCTGGCTGGCTTCGTTATAGGAACGGC 240
 QY 81 SerThrAsnGlyGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 DB 241 TCCACTAATAAAGGAATGCGCTCAGCGGAATCTAGTCTTCCAACTATCATGTGATAGC 300
 QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 DB 301 GGTGGGGGACTTGAGGACTACCTTCGATCTGCAACCTTATTCAGCAACATACAGT 360
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
 DB 361 GCTGGTGCCAGATTCATACAACTCCTGGGAGCAGCAGTGAATGGGCTTTACACAACA 420
 QY 141 AspSerArgAsnValaAspTyrValaArgLysAsnAspMetThrIleLeuPheAlaAla 160
 DB 421 GATTCAGAAATGGTATGACTATGTCCGCAAAATGATATGACATCTTTTCGCTGCC 480
 QY 161 GlyAsnGlyGlyProAsnGlyGlyTrpIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 DB 481 GCGAATGAAGCCGGAACCGGGAACCACTCAGTGCACAGGACAGCAGTAAAAATGCAATA 540
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAsn 200
 DB 541 ACAGTCGGAGCTACGGAACCTCCGCCCAAGCTTTGGGTCTTATCGGACATATCAAC 600
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 DB 601 CATGTGCACAGTCTCTTCCTTCACGTGGACCGCAAAAGGATGGAGTCAACCGGATGTC 660
 QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
 DB 661 ATGGACCGGGAACGTTCTATCATCAGCAAGATCTTCTTGCACCGGATTCCTCCCTTC 720
 QY 241 TrpAlaAsnHisAspSerIlystYrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260

DB 721 TGGCGAACCATGCACAGTAATATGATCATATGCTGGTGAACGTCCTCATGCTACCCGATC 780
 QY 261 ValAlaGlyAsnValaAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 DB 781 GTTCTGGAAACGTGCACAGCTTCGTGAGCATTTTGTGAAAAACAGAGGCTACACCA 840
 QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 DB 841 AAGCCTTCTCTATTAAAGCGGCACCTGATGCGGTGCAGCTGACATCGGCTTGCTAC 900
 QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 DB 901 CCGAACGGTAACAGGATGGGACGAGTGACATGATGATAAATCCCTGAACGTTGCTAT 960
 QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTySerPheThrAlaThr 340
 DB 961 GTGAACGAGTCCAGTTCTCTATCCACGACGCAAAAGGACGCTACTCGTTTACTGTACT 1020
 QY 341 AlaGlyIysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 DB 1021 GCCGGCAAGCCTTTGAAATCTCCCTGGTATGCTGTGATGCCCCCTCGGACCAACTGCT 1080
 QY 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 DB 1081 TCCGTAAAGCTTGTCAATGATCTGGACCTTGCTATCCGCTCCAAATGGCACAGTAT 1140
 QY 381 ValGlyAsnAspPheThrSerProTyAsnAspAsnTrpAspGlyArgAsnValGlu 400
 DB 1141 GTAGGAATGACATTCTTCGCCATCAATGATTAACCTGGGATGGCCGCAATACGAGAA 1200
 QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyThrIleGluValGlnAlaTyrAsn 420
 DB 1201 AATGATTATTATATGCACCAAAAGCGGAGCTATACATGAGGTACAGGCTTATAAC 1260
 QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 DB 1261 GTACCGGTTGCACACAGACCTTCTCGTTGGCAATTGTGAAT 1302
 RESULT 2
 US-10-090-624-11
 ; Sequence 11, Application US/10090624
 ; Publication No. US2002013235A1
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAKURA, Hikaru
 ; APPLICANT: MORISHITA, Mio
 ; APPLICANT: SHIMOJO, Tomoko
 ; APPLICANT: ASADA, Kiyozo
 ; APPLICANT: KATO, Ikunoshin
 ; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
 ; FILE REFERENCE: TAKAKURA=6
 ; CURRENT APPLICATION NUMBER: US/10/090,624
 ; PRIOR FILING DATE: 2002-03-06
 ; PRIOR APPLICATION NUMBER: 09/445,472
 ; PRIOR FILING DATE: 1999-12-06
 ; PRIOR APPLICATION NUMBER: 151969/1997
 ; PRIOR FILING DATE: 1997-06-10
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 11
 ; LENGTH: 1977
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Synthetic
 ; US-10-090-624-11
 Alignment Scores:
 Pred. No.: 2,62e-40 Length: 1977
 Score: 452.50 Matches: 138
 Percent Similarity: 44.66% Conservative: 67
 Best Local Similarity: 30.07% Mismatches: 153
 Query Match: 20.14% Indels: 101
 DB: 13 Gaps: 18

Db 171 TAGGAGTTATCCATACGATGACATGACATGGAATCTATAGCTTCAATAGCAGCTGG 230
Qy 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 231 TACTGCAGCACAAGTAATGCAAGTCAAGGGAATGGCTCCAGAGACTAGCTGGCGGG 290
Qy 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyLeuProSerAsnLeuGl 112
Db 291 AATTAAAGGTTCTAGGTGCGATGGTCTGGAAGCATATCTACTATATAATTAAGGAGTGA 350
Qy 112 nThrLeuPheSerGlnAlaTySerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 351 GTGGCGCGTTGATAACAAGATAAGTACGGAATTAAGGTCTAATCTCTCTGTTGTC 410
Qy 132 a-----AlaValAsnGlyAlaTyThrTh 140
Db 411 AGCCAGAGCTCAGATGCTAGTACGCTCTAAGTACGCTGTTAATGCGAGGTGGATGC 470
Qy 140 rAspSerArgAsnValAspAspTyValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 471 T-----GGATTAGTTGTTGGTTGCCGC 494
Qy 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIl 180
Db 495 TGAACAAGTGGACCTAACAGTATACAAATCGGTTCTCAGCAGCTGCACCAAGTTAT 554
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAs 200
Db 555 TACAGTTGGAGCC-----GTTGACAAGATGA 581
Qy 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 582 TGTTTAACACAGCTTCTCAAGCAGAGGGCCAACTCGACAGCGAGCTTAAGCTGAGGT 641
Qy 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 642 TGTGTCTCCAGGAATCGGATAATTGCTGCCAGACAAAGT-----GGAAGTACGAT 692
Qy 240 eTrpAlaAsnHisAspSerLysTyAlaTyMetGlyThrSerMetAlaThrProIl 260
Db 693 GGGTCAACCAATTAATGACTATTACACAGCAGCTCCTGGGACATCAATGGCAACTCTCA 752
Qy 260 eValAlaGlyAsnValAlaGlnLeuArgLysPheValLysAsnArgGlyIleThrPr 280
Db 753 CGTAGCTGTATGCGAGCCCTCTGCTCCAA-----GCACACCC 791
Qy 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 792 GAGCTGGACTCCAGACAAAGTAAACACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 851
Qy 294 -----AlaAspIleGlyLeuGlyTyProAsnGlyAsnGlnGlyTyTrpGlyValVa 310
Db 852 AGATGAATAGCCGATATAGCTACGGTGCA-----GGTAGGGT 890
Qy 310 lThrLeuAspLysSerLeuAsnValAlaTyValAsnGluSerSerSerLeuSerThrSe 330
Db 891 TAATGCATCAAGGCTATAAAC-----TACGATAACTATGCAAGAGTAGTGTCACTGG 944
Qy 330 rGlnLysAla-----ThrTySerPheThrAlaThrAlaGlyLysProLe 345
Db 945 ATATGTTGCCAACAAAGGAGCAAACTCCACGTTCTGTTATAGGGAGCTTCCTCGT 1004
Qy 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1005 AACTGCCACATTATACCTGGGACAAATGCCAAT----- 1035
Qy 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyValGlyAsnAspPh 385
Db 1036 -AGCGACCTTGATCTTACTCTTACCATCCCAATGGAAACCCAG---GTTGACTACTCTTA 1091
Qy 385 eThrSerProTyAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405

Db 1092 CACCGCTACTAT-----GGATTGAAAAGGTTGGTTATTA 1127
Qy 405 nAlaProGlnSerGlyThrTyThrIleGluValGlnAlaTyAsnValProValGlyPr 425
Db 1128 CAACCCACTGATGGACATGACATTAAGTTTGAAGTACAGC-----GGAAG 1178
Qy 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1179 TGCAAACTATCAAGTAGATGCTGTAAGT 1206
RESULT 4
US-10-090-624-15
; Sequence 15, Application US/10090624
; Publication No. US200203235A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-15
Alignment Scores:
Pred. No.: 2,82e-36 Length: 1962
Score: 416.50 Matches: 141
Percent Similarity: 42.55% Conservative: 59
Best Local Similarity: 30.00% Mismatches: 148
Query Match: 18.54% Indels: 122
DB: 13 Gaps: 20
US-09-985-689A-1 (1-434) x US-10-090-624-15 (1-1962)
Qy 12 ValAlaGlnSerSerTyrgly-LeuTy-----GlyGlnGlyGl 24
Db 420 GTCTGCAGCTCAAGTTATGCGCACTTACGTTGGAACTTGGGATATGATGGTTCTGGAAT 479
Qy 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGl 44
Db 480 CACAATAGGATTAATTTGACACTGGATTGAC-----GCTTCTCATCC 521
Qy 44 uAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAAAGGAAAAGTA-----ATTGGGTGGTATGATTTTGTCAATCG 566
Qy 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGl 78
Db 567 TAGAGTTATCCATAGTACCATCGCATGGAACTCATGTAGCTTCAATAGCAGCTGG 626
Qy 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGGCAAGTACAAAGGAATGGCTCCAGGAGCTAAGCTGCGGG 686
Qy 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGl 112
Db 687 AATTAAAGTTCTAGGTGCGCATGTTCTTCTGGAAACCATATCTACTATTAATTAAGGAGTTGA 746
Qy 112 nThrLeuPheSerGlnAlaTySerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132

```
Db 747 GTGGGCGCTGATACAAAGATAGTACGGAATTAAGTCAATTAATCTTCTTGGTTC 806
Qy 132 a-----AlaValAsnGlyAlaValThrTh 140
Db 807 AAGCCAGAGCTCAGATGCTAGCTCTAAGTCAAGTCTGTTAATGACGCGTGGATGC 866
Qy 140 rAspSerArgAsnValAspAspThrValArgLysAsnMetThrIleLeuPheAla 160
Db 867 T-----GGATTAGTTGTTGTTGTTGCCGC 890
Qy 160 aGlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAla 180
Db 891 TGGAAACAGTGGACCTTCAAGATACATCGCTCTCCAGCAGCTGCAAGCAAGTTAT 950
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAs 200
Db 951 TACAGTTGGAGCC-----GTTGACAAAGTATGA 977
Qy 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyValArgIleLysProAspVa 220
Db 978 TGTATTAAACAAGCTTCTCAGCAGAGGGCCAACTGCAGACGGGAGGCTTAAGCTGAGGT 1037
Qy 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 1038 TGTGTCTCCAGAACTGGATAATTGCTGCAGAGCAAGT-----GGAACTAGCAT 1088
Qy 240 eTTPAlaAsnHisAspSerLysTyAlaTyMetGlyThrSerMetAlaThrPro 260
Db 1089 GGGTCAACCAATTAATGACTATTACAGCAGAGCTCTCGGACATCAATGGCACTCTCTCA 1148
Qy 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgIleThrPr 280
Db 1149 CGTAGCTGGTATTGCAGCCCTCTGCTCCAA-----GCACACCC 1187
Qy 280 oLys-----ProSerLeuLeuLysAlaAlaLeuAlaGlyAla----- 293
Db 1188 GAGCTGACTCCAGACAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 1247
Qy 294 -----AlaAspIleGlyLeuGlyTyProAsnGlyAsnGlnGlyTrpGlyArgVa 310
Db 1248 AGATGAAATAGCGATATAGCTACGCTGCA-----GGTAGGTT 1286
Qy 310 lThrLeuAspLysSerLeuAsnValAlaTyValAsnGluSerSerSerLeuSerThrSe 330
Db 1287 TAATGCATACAAGCTATNAAC-----TACGATAACTATGCAAAAGCTAGTTCTACTGG 1340
Qy 330 rGlnLysAla-----ThrTySerPheThrAlaThrAlaGlyLysProLe 345
Db 1341 ATATGTTGCCAACAAAGCGACCAACTCACCAGTTCGTTATTAGCGGAGCTTCGTTCTGT 1400
Qy 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1401 AACTGCCACATATTACTGGGACAAATGCAAT----- 1431
Qy 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyValGlyAsnAspPh 385
Db 1432 -AGGACCTTGATCTTACTCTACGATCCCAATGGAACAG-----GTTGACTACTCTTA 1487
Qy 385 eThrSerProTyAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1488 CACCGCTACTAT-----GGATTGAAAAGGTTGGTTATTA 1523
Qy 405 nAlaProGlnSerGlyThrTyThrIleGluValGlnAlaTyAsnValProValGlyPr 425
Db 1524 CAACCCAACTGATGGAACATGGCAATTAAGTTGTAAGCTACAGC-----GGAAG 1574
Qy 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1575 TGCAAACTATCAAGTAGATGTGGTAACT 1602
RESULT 5
US-10-156-761-5701
```

```
; Sequence 5701, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5701
; LENGTH: 3624
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3624)
US-10-156-761-5701

Alignment Scores:
Pred. No.: 6,23e-30 Length: 3624
Score: 363.50 Matches: 128
Percent Similarity: 45.50% Conservative: 54
Best Local Similarity: 32.00% Mismatches: 145
Query Match: 16.18% Indels: 73
DB: 14 Gaps: 18

US-09-985-689A-1 (1-434) x US-10-156-761-5701 (1-3624)
Qy 8 VallysAlaAspValAlaGlnSerSer-----Tyr 17
Db 556 GTGAGGCGGACGATCGCCGAGCAAGCGCAGATCGTACGCGGGCGCGGTGGAGCGCC 615
Qy 18 GlyLeuTyGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 616 GGGCTCACGGCGGCGGCTGTCGCTGCGCTGCTGACACCGGCGTGCACACC----- 669
Qy 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAlaGly 57
Db 670 -----ACTCACCCGACCTCGCGCGGGGTGTCCCGGAGCAAGAGCTTCATC 717
Qy 58 ArgThrAsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 718 GACGGGGAGGAGGTGCGCGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 777
Qy 78 GlyAsnGlySer-----ThrAsnLysGlyMetAlaProGlnAlaAsnLeuVal 93
Db 778 GGCAGCGCGCGCTCCGACGCGGACGCGGCGCGCTGCGCGCGCGCGCGCGCGCGCGCG 837
Qy 94 PheGlnSerIleMet---AspSerGlyGlyLeuGlyLeuGlyLeuProSerAsnLeuGln 112
Db 838 CTCGCAAAAGTGCTCAGCAGCAGCGCGCGGGA-----AGGAGTCCCGAG 882
Qy 113 ThrLeuPheSerGlnAlaTySerAla-----GlyAlaArgIleHisThrAsnSer 129
Db 883 ATCATCGCGGCGATGGATGGCGCGGCGGACGCGGCGGACGCGTGCAGGATCGTCTCGATGAGC 942
Qy 130 TrpGlyAla-----AlaValAsnGlyAlaTyThrThrAsp-----SerArgAsnVal 145
Db 943 CTCGGATCGACCGAGCGGACGCGG-----ACCGACCCCATGGCGGCGCGGCGGCG 993
Qy 146 AspAspTyValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlyPro 165
Db 994 GACACCTCTCGAGGAGACCGCGCGCCCTCTTCTGCTGCGCGGCGGGAACACCGGTCGCC 1053
```


QY 285 LeuLysAlaAala-----LeuLleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsn 302
 Db 6919614 ATGTCACGTCGAGCAACTCGACCGCTCCGATATACATTTGGGGCGGGTGGGTGAGT 6919673
 QY 303 GlyAsnGlnGlyTyrGly---ArgValThrLeuAspLysSerLeuAsnValAlaTyr--- 320
 Db 6919674 GTGCGGAGCGGTCGCGCGCGTCAACGCGAGCGGCAAGCGGCGGCTTCCAC 6919733
 QY 321 -----ValAsnGlnSerSerSerLeuSerThrSerGln 331
 Db 6919734 CGCTGGCCCATGACCGCGATGACCGCTCAGAGAGCGGTCACTACTCCATCTCCTCC 6919793
 QY 332 LysAlaThrTyrSerPheThr---AlaThrAlaGlyLysProLeuLysIleSerLeuVal 350
 Db 6919794 GACACGACGTCGTGAGTTCGCGTCCGCGGCGCGCGCGCGGCGGTGTCGCCACCGCTC 6919853
 QY 351 TrpSerAspAla-----ProAlaSerThrThrAlaSerValThrLeuValAsnAsp 367
 Db 6919854 GCCGACCGCGCACTACCGGTGCGCGCGCCGCCACCGCGCGCCACCGGTGACCGCGGAC 6919913

RESULT 7
 US-10-156-761-3306
 ; Sequence 3306, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 3306
 ; LENGTH: 3417
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(3417)
 US-10-156-761-3306

Alignment Scores:
 Pred. No.: 4,6e-28 Length: 3417
 Score: 346.50 Matches: 137
 Percent Similarity: 43.81% Conservative: 54
 Best Local Similarity: 31.42% Mismatches: 177
 Query Match: 15.42% Indels: 70
 DB: 14 Gaps: 15

US-09-985-689A-1 (1-434) x US-10-156-761-3306 (1-3417)

QY 18 GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
 Db 745 GGGTACGACGCGAAGGGGTCAGATCGCGTCTCGGACACCGGTGTCGAC-----795
 QY 38 AsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
 Db 796 -----GCCACCCACCGGACCTCAAGACACAGGTCGGCGGAGTCACGAACCTTCTCC 846
 QY 58 ArgThrAsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
 Db 847 GCCCG 906
 QY 78 GlyAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 93

Db 907 GGCACCGCGCAAGTCCAAACGGCAAGTACAAAGGTGTCCGCGCGGCGCGAGATCCCTC 966
 QY 94 PheGlnSerIleMetAspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThr 113
 Db 967 AACGCGAAGGTCTCGACGACACCGGC-----TCCGGCGACGACTCCGCGCATCTGGCC 1020
 QY 114 LeuPheSerGlnAlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAl 133
 Db 1021 GCGATGGAGTGGGGCGGCGGACGAGGCGCGGAGCTGTCTCAACTGAGCTTGGCGGCGGC 1080
 QY 133 aValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAs 153
 Db 1081 GACACCCCGAGATCGACCGCTGGAAGCGAG-GTCAACAGCTCTCCGAGGAGAGGG 1139
 QY 153 pMetThrIleLeuPheAlaAlaGlyAsnGlyGlyProAsnGly---GlyThrIleSerAl 172
 Db 1140 CATCTCTTCGCGATCGCGCGGCAAGGAGGAGTTCGGGAGGAGACCATCGCTC 1199
 QY 172 aProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPh 192
 Db 1200 CCGGCGACGCGCGCGCGCTCACCGTCCGCGGC-----1236
 QY 192 eGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrIly 212
 Db 1237 -----GTGAACGACGACGACGAGCTGGGTCTCTTCAGCGCGCGCGCGCTC 1286
 QY 212 aAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSe 232
 Db 1287 GAGCGGCGCATCAAGCCGACGTCACCGCACCGCGGTGGAGATCACCGCGCGC-----1341
 QY 232 rSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLys-----TyrAl 249
 Db 1342 -----GCGGCGCGCGGCGCGCTCATCGACGAGGAGTCCGCGAGCGCGCGCTACCT 1397
 QY 249 aTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuAr 269
 Db 1398 CACCATCTCCGCTACGTGCGAGCCCGCATGTCCGCGGCGCGCGCGCGCTCTCAA 1457
 QY 269 gGluHisPheValLysAsnArgGlyIleThrProLysProSerLeu-----LeuLysAl 287
 Db 1458 CGAGCGAC-----CCCACTGCTGTTCCGCGAGTCAAGGG 1496
 QY 287 aAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTr 307
 Db 1497 CCGCTGACCGGCTCCGCAAG-----GGCGGCAAGTACACGCGCTTCAGCAGGCTC 1550
 QY 307 pGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlnSerSerLe 327
 Db 1551 GGGCGGTATCCGCTCGACAGGCGATCAAGCAGTCCGTGATGCCACCGCGACTCGGT 1610
 QY 327 uSerThrSerGlnLysAlaThrTyrSerPhe-ThrAlaThrAlaGlyLysProLeuLysI 347
 Db 1611 G-----AGCTTCGGCATCCAGCAGTGGCGCGCACACCGACGA 1646
 QY 347 leSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuValAsnA 367
 Db 1647 CAAGCGGTCAACCGCAGAGCTCACCTACCGCAACTCCGCGCGAGTGACGTC-ACGCTGA 1705
 QY 367 pLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPhe---- 385
 Db 1706 ACCTCGCTCGACGCGCACCAACCCCAAGGGCGTGGCGCTCCGCTCGGGCTTCTCAAGC 1765
 QY 386 -----ThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValG 400
 Db 1766 TCGGCGCGACGAAGGTACGGTCCCGCGGCG-----GGCAAGGCTCGGTG 1813
 QY 400 lu-----AsnValPheIleAsnAlaProGlnSerGlyThrThrIleGluValG 417
 Db 1814 ACTTCAGGTCACACGAGCTGGCGGCGCACCGACGCGCGGTACTCGCGCTACGTGA 1873
 QY 417 lAlaTyrAsnValProValGlyProGlnThrPheSerLeuAla 431


```
; PRIOR APPLICATION NUMBER: US 60/278,935
; PRIOR FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 214
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 1
; LENGTH: 135638
; TYPE: DNA
; ORGANISM: Streptomyces atroolivaceus
US-10-314-657-1

Alignment Scores:
Pred. No.: 1,056-24 Length: 135638
Score: 338.00 Matches: 127
Percent Similarity: 43.06% Conservative: 59
Best Local Similarity: 29.40% Mismatches: 144
Query Match: 15.04% Indels: 102
DB: 17 Gaps: 17

US-09-985-689A-1 (1-434) x US-10-314-657-1 (1-135638)

QY 6 GlyIleValIysAlaAspValAlaGlnSer----- 15
Db 8067 GGCAGGGTGAGGCGGATCTGGCGGACTCCACCGCCAGATCGGCGCGCAGAGGTATGG 8126

QY 16 SerTyGlyLeuTyrglyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThr 35
Db 8127 GCGAGGGCCACACCGCCAGGACGTGAAGGTGCGATGCTGCACAGCGCGCGGACACC 8186

QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIle-----ThrAlaLeu 53
Db 8187 -----GAAACCCCGACCTGGTGGGCGAGGTGTCGACAGCGCCAGC 8228

QY 54 TyrAlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAla 73
Db 8229 TTCGTCCCGCGGAGGACGACATCGCC---GACTACAAACCGCCAGCGCAGCGCTCGCC 8285

QY 74 GlySerValLeuGlyAsnGlySerThrAsp-----LysGlyMetAlaProGln 89
Db 8286 TCAGACCATGTGCGACCGGCGGCGGCTCGACGCGGAGGCGGGGTGTCGCTCCGCG 8345

QY 90 AlaAsnLeuValPheGlnSerIleMetAspSerGlyGlyGly----- 103
Db 8346 GCGCGGTGTGCGTGGCAGGAGTGTCACTCCAGGCGGCGGCGGCGGAGGATCGTGGATC 8405

QY 104 LeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrrSerAlaGlyAla 123
Db 8406 ATCCGCGGCATGGAG-----TGGCGCCCGCGCGGCGGCGGAGGAGGCG 8444

QY 124 ArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrrThrAsp----- 141
Db 8445 AGGATCATCAGCATGAGCCTGGGC-----GGCGCGGTGACAAAGAACCGCGATG 8495

QY 142 SerArgAsnValAspAspTyrrValArgLysAsnAspMetThrIleLeuPheAlaAlaGly 161
Db 8496 AGCCAGGCGGTGCGAGCACTCAGCCACGACGCGGCGGCGGTGTCGTGTGATCGCGGCGGC 8555

QY 162 AsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 181
Db 8556 AACCGCGCGCCGAC-----TCCATCAGGAGCGCGGTGCGGAGCTCCGCGCTGACC 8609

QY 182 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrrAlaAspAsnIleAsnHis 201
Db 8610 GTCGCGGCC-----GTCGACTCCACCGACACG 8636

QY 202 ValAlaGlnPheSerSerArgGlyProThrLys---AspGlyArgIleLysProAspVal 220
Db 8637 CTCGCGGACTTCTCCAGCGGCGCGCGGTGACGCGGCGGCGGCTGAAGCGCGGAGATC 8696

QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 8697 ACCGCGCGCGCGTCGATCGTCGCGCGCGCTCGCACTACAAAGCGCGGCTCCGCGTAC 8756

QY 241 TrpAlaAsnHisAspSerLysTyrrAlaTyrrMetGlyGlyThrSerMetAlaThrProIle 260
Db 8757 -----TACACCAACGATGAGCGGCACTCGATGGCGAGCGCGCAC 8795

QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIysAsnArgGlyIleThrPro 280
Db 8796 GTCGCGGTGTGCGCGGCTCTCGCGCGGAGCACCCCGACTGGACGGGACCCAGCTC 8855

QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyrr 300
Db 8856 AAGGAGGCACTGGTCAGCAGCGCCAGGCAAGCGCGCG-----TAC 8897

QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrr 320
Db 8898 ACCCGGTACAGGCGGCGCGCGCGCTCGACGCGCGCGCGCGCGCGCGCGCGCGCG 8957

QY 321 ValAsnGluSerSerSerLeuSer-----ThrSer 330
Db 8958 TTCGCCACCGACCGCGCTACTCCGGTTCACACGTCGGCCCCCGAAGCGCGGAGACC 9017

QY 331 GlnLysAlaThrTyrrSerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuVal 350
Db 9018 GATGTCGCGAGCGGTGACATCACCAACGTCGCGGACGCCCGCGTCAACCTCGGCC 9077

QY 351 TrpSer-----AspAlaPro 355
Db 9078 GTCACCGCACCGTCCCGCGGGTGTTCAGCTCTCCGAGGACCATGTCACCGTGGCC 9137

QY 356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAsp----- 369
Db 9138 GCGCACGCGCACCGCGCGGTCCACTGACCGCGCTCTGGACAGCTGGCGGCGGACGAG 9197

QY 370 -----LeuValIleThrAlaProAsnGlyThr 378
Db 9198 TCGGTACGCGCGGTGATCACCGTACGCGCGGCGGACG 9233

RESULT 10
US-09-974-300-1934
; Sequence 1934, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; INVENTOR: Clausen, Ib Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; FILE REFERENCE: 10085-500-US
; CURRENT FILING DATE: 2001-10-05
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1934
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1934

Alignment Scores:
Pred. No.: 4,128-27 Length: 1329
Score: 332.50 Matches: 104
Percent Similarity: 43.53% Conservative: 44
Best Local Similarity: 30.59% Mismatches: 109
Query Match: 14.80% Indels: 83
DB: 9 Gaps: 13

US-09-985-689A-1 (1-434) x US-09-974-300-1934 (1-1329)

QY 11 AspValAlaGlnSerSerTyrrGlyLeuTyrrGlyGlnIleValAlaValAlaAsp 30
Db 406 GAAGTGGTCAGAAACATCAGACGCTGACGCAAGGAGTGACAGTCGCTGTCATTGAT 465
```



```
QY 188 LeuArgProSerPheGlySerTyrAlaAsp----- 197
Db 1669 GTCCCAATTAACGTTGAGTTATTTCCCAAGCACTTGGATATCCTGATTACTATGGA 1728
QY 198 -----AsnIleAsnHieValAlaGlnPheSerSerArgGly 209
Db 1729 TTCATTACTTCCCGCCGTACACAAAGTT--AGAATAGCATTTCTTCAAGCAGAGGG 1785
QY 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229
Db 1786 CCGAGAATAGATGTTGAATAAAACCAATGATAGTGGCTCCAGGTTACGGAATTTACTCA 1845
QY 230 AlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHieAspSerLysTyrAla 249
Db 1846 TCCCTGCCGATGGATTGGCGGAGCTGCTTC----- 1878
QY 250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArg 269
Db 1879 ---ATGCTGGAACCTCGATGGCTACTCCACATGTCCAGCGGTGCTGTCATCTCTCAT 1935
QY 270 GluHiePheValLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu 289
Db 1936 AGCGGG---GCAAGCGCGAGGGAATATACTACATCCAGATATATTAAGAGGTTCTT 1992
QY 290 IleAlaGlyAlaAlaAspIle-----GlyLeuGlyTyrProAsnGly 303
Db 1993 GAGAGCGGTGCAACCTGGCTTGAGGAGATCCATATCTGGGCGAGAGTACACTGAGCTT 2052
QY 304 AsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlu 323
Db 2053 GACCAAGGTGATGCTGTTAAGCTTCCAGTATCCAAAGTCTGGGAATCTTAAGGCTATAAC 2112
QY 324 SerSerSerLeuSerThrSerGln-LysAlaThrTyrSerPheThrAlaThrAlaGlyLys 343
Db 2113 GGCACCACTCTCCAAATGTTGATCTACTGGCGAGACAGTCTTACAGCGACT----- 2164
QY 343 sProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAlaSerValTh 363
Db 2165 ---TTCGGAGTACTTGGGTGGAGCTTATAAGAGTCTCTACGCAAGGAACT----- 2215
QY 363 rLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThr--GlnTyrValGly 382
Db 2216 -----CTATACCTGACATTTTGAGATCTATGCAACTGAGCCATGGATTAAAGCCTTTTGT 2316
QY 383 Asn-----AspPheThrSerProTyrAsnAspAsnTrp----- 393
Db 2257 GACACGGAGTACAGAACTTTTGAGATCTATGCAACTGAGCCATGGATTAAAGCCTTTTGT 2316
QY 394 AspGly-----ArgAsnAsnValGluAsnValPhe-----Ile 404
Db 2317 AGTGAAGTGTAAATCTAGAGAACATACCAGTTTGTCTTATAGGTGAATATGATGTA 2376
QY 405 AsnAlaProGlnSerGlyThrTyr----- 412
Db 2377 GAGGCTCTGAGCGAGTCTCTATGTTGAAGGATAATCATTTGATATCCAAACGCCA 2436
QY 413 ThrIleGluValGlnAlaTyrAsnValProValGlyProGlnThrPheSer 429
Db 2437 GTTATTGAAGACGAGATCTTGAACAAATGTTATTCCCGAAGATTCTACT 2487

RESULT 12
US-10-084-846A-113
; Sequence 113, Application US/10084846A
; Publication No. US2004006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084, 846A
; CURRENT FILING DATE: 2003-02-25
```

```
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 113
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: avix16 dna: partial sequence of coding strand 1; nucleotide 1
; OTHER INFORMATION: corresponds to nucleotide 56,198 of coding strand 1.
US-10-084-846A-113

Alignment Scores:
Pred. No.: 4,32e-24 Length: 1560
Score: 306.50 Matches: 116
Percent Similarity: 39.20% Conservative: 51
Best Local Similarity: 27.23% Mismatches: 154
Query Match: 13.64% Indels: 105
DB: 15 Gaps: 19

US-09-985-689A-1 (1-434) x US-10-084-846A-113 (1-1560)
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 451 GGGCAGGAGTACGGCGGTACTCATCGACACACCGCGCTC-----CGC 492
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr 59
Db 493 ATCACCCACAGCACTTCGGCGCGCGCTCTCCACGCTACGCGTACGCGCATCGACACGAC 552
QY 60 AsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 553 AACACCGCCAGGACGGCCACGCGCACGCGCACGCGCGCGCACGCGTCCGCGGCAAC 612
QY 80 GlySerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 99
Db 613 GCC-----TACGGCGTCCCAAGAGCCCAAGATCGTAGCGCTCCGCGTCTGAAC 663
QY 100 SerGlyGly-----GlyLeuGlyGlyLeuProSerAsn 110
Db 664 AACTCCGCGCAGGACACACCGCCCGCTCGTCCGCGCATCGTGGTCCGCGCGAAC 723
QY 111 LeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrp 130
Db 724 -----GCCGTCAAGCGCGCGCTCGCCCAACATGCTCCCTC 756
QY 131 GlyAlaAlaValAsnGlyValaTyrThrThrAspSerArgAsnValAspAspTyrValArg 150
Db 757 GCGCGCGCGCGCACCGCCCTCGACACGCGCGCTACGCAAC-----GCCATG 804
QY 151 LysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIle 170
Db 805 GCCTCCGCGCTCACCTTCGCGTGGCGCGCGCGCACGAGTGCACCAACGCTCCACGAGG 864
QY 171 SerAlaProGlyThrAlaAlaAsnAlaIleThrValGlyAlaThr-----GluAsnLeu 188
Db 865 TCA---CCCGCAGCGGTCCAGGCGCATCAGTTCGGCGCGCAGCAGCAGCTCGGACGCC 921
QY 189 ArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArg 208
Db 922 AAGCGCGGTACTCCCAACTACGGCTCCGCTC----- 954
QY 209 GlyProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeu 228
Db 955 -----GACCTTCGCGCGCGGTTCGTCATCATCACC 984
QY 229 SerAlaArgSerSerLeuAlaProAspSerPheTrpAlaAsnHisAspSerLysTyr 248
Db 985 TCGGCC-----TGGAACCTCAAGCGAGTCTCGCGGACC 1014
```

QY 249 AlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu 268
Db 1015 AACACCACTCCGGTACGTGATGGCGACCCCGACAGTGGCGGGCGCGGCTC--- 1071
QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeu 285
Db 1072 -----CACCTCGCGCCCAACCCCTCGGCCACCCCGCTCCAGGTCGCGGCTGACG 1125
QY 286 LysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGln 305
Db 1126 TCCGCGCCACACACCGCGGTGTCTACCAACCCCGCGGCGGCTCGCCAC- 1176
QY 306 GlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaLysValAsnGlu----- 323
Db 1177 -----CGGCTCCTGTACTCGTGGCGCGCGGCAC 1203
QY 324 -----SerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAla 339
Db 1204 GACACACCTCCGGGCGCGGCTTCGAGAACACCGGTGACTACAGTACGCGCAACTCC 1263
QY 340 ThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThr 359
Db 1264 ACCTGCGAGTCCCGGTACGCTCCCGGCTCCCGCAACCGCGGCTCGGCTCGCC 1323
QY 360 AlaSerVal-----ThrLeuValAsnAspLeuAspLeuValIleThrAlaPro 375
Db 1324 GTAGAGGTCCACATCGTCCACACGATACGCGGACCTCCAGGTCAGCTGATCGCCGCC 1383
QY 376 AsnGlyThrGlnTyrValGlyAsnAspPhe---ThrSerProTyrAsnAspAsnTrpAsp 394
Db 1384 GACGCGACGGCGGTACACGCTCAAGTCGATCGACCGCACCGCGCGGCTCGGAC- 1434
QY 395 GlyArgAsnAsnValGluAsnValPhe---IleAsnAlaProGln-----SerGly 410
Db 1435 -----ACATCAACACACACGATCGGTGACCGCTCTCCGAGGCGCGCCACACGCG 1485
QY 411 ThrTyrThrIleGluVal 416
Db 1486 ACGTGGAACTCGGGTG 1503

RESULT 13

US-10-084-846A-1

; Sequence 1, Application US/10084846A

; Publication No. US20040006026A1

; GENERAL INFORMATION:

; APPLICANT: WEITNAUER, GABRIELE

; APPLICANT: TREFFER, AXEL

; APPLICANT: MUHLWEG, AGNES

; APPLICANT: BECHTOLD, ANDREAS

; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES

; FILE REFERENCE: 1974-005

; CURRENT APPLICATION NUMBER: US/10/084,846A

; CURRENT FILING DATE: 2003-02-25

; PRIOR APPLICATION NUMBER: PCT/EP01/09815

; PRIOR FILING DATE: 2001-08-24

; PRIOR FILING DATE: 2001-02-25

; NUMBER OF SEQ ID NOS: 120

; SOFTWARE: Patent in Ver. 3.2

; SEQ ID NO 1

; LENGTH: 59816

; TYPE: DNA

; ORGANISM: Streptomyces viridochromogenes

US-10-084-846A-1

Alignment Scores:

Pred. No.:	1,046-21	Length:	59816
Score:	306.50	Matches:	116
Percent Similarity:	39.20%	Conservative:	51
Best Local Similarity:	27.23%	Mismatches:	154
Query Match:	13.64%	Indels:	105
DB:	15	Gaps:	19

US-09-985-689A-1 (1-434) x US-10-084-846A-1 (1-59816)

QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 56648 GGGCAGGAGTACACGGGTACGTATCGACACCGGCGTC-----CGC 56689
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr 59
Db 56690 ATCACCACACGAGCTTCGGCGCGCGGCTCTCTACGCGTACGAGCCATCGAACACGAC 56749
QY 60 AsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 56750 AACACCGCCACGACGGCCACCGCCACGCGACGACGTGGCGCGGACCGGTCCCGCGCAC 56809
QY 80 GlySerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 99
Db 56810 GCC-----TACGGCGTCGCCAAGAACGACAGTCTAGGGGTCCGGCTGTGTAAC 56860
QY 100 SerGlyGly-----GlyLeuGlyGlyLeuProSerAsn 110
Db 56861 AACTCCGGCAGGGCACACCCCGCAGGTCTGCGCGGATCGACTGGTCCCGCGGAC 56920
QY 111 LeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrp 130
Db 56921 -----GCCGTCAAGCGCGCTCGCCCAACATGTCTCCCTC 56953
QY 131 GlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArg 150
Db 56954 GCGCGCGCGCCACACGCGCTCGACACGCGCTACGCAAC-----GCCATG 57001
QY 151 LysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIle 170
Db 57002 GCCTCCGGCGTCACTTCGCGTGGCGCGGCAACGAGTCCGACCAACGCTCCACGAGG 57061
QY 171 SerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeu 188
Db 57062 TCA---CCCGCAGCGGTCCACGAGGCCATCAGGTGGCGCGGCGCACCAAGTCCGACGCG 57118
QY 189 ArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArg 208
Db 57119 AAGCGCGGTACTCCAACTACGCTCGGCTCC----- 57151
QY 209 GlyProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeu 228
Db 57152 -----GACCTCTTCGCCCCCGGTTCGTCATCACC 57181
QY 229 SerAlaArgSerSerLeuAlaProAspSerPheTrpAlaAsnHisAspSerLysTyr 248
Db 57182 TCGGCC-----TGAACTCAAGCGACTCGGCGACC 57211
QY 249 AlaTyrMetGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu 268
Db 57212 AACACCATCTCCGGTACGTGATGGCGACCCCGCGACGTGGCGCGCGCGCTC--- 57268
QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeu 285
Db 57269 -----CACCTCCCGCCAAACCCCTCGGCCACCCCGTCCCGAGTCCCGCGGCTGACG 57322
QY 286 LysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGln 305
Db 57323 TCCGCGCGCACACCGCGGTCTGTCACCAACCCCGCGGCGGTTCGCCCAAC----- 57373
QY 306 GlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlu----- 323
Db 57374 -----CGGCTCTGTACTCGCGCGGCGGAC 57400
QY 324 -----SerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAla 339
Db 57401 GACCACTCCCGCGCGGCTTCGAGAACACCGCGGTGACTACACGATCAGCACAACCTCC 57460
QY 340 ThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThr 359
Db 57461 ACGGTGAGTCCCGGTGACGGTCTCCGGCGTCTCCGGCAACCGCGGCTCGGCCCTCGCC 57520

QY 360 AlaserVal-----ThrLeuValAsnAspLeuAspLeuValIleThrAlaPro 375
 Db 57521 GTAGAGTCCACATCGTCCACACGTATCGCGGACCTCCAGTCCAGCTGATCGCCCC 57580
 QY 376 AsnGlyThrClnTyValGlyAsnAspPhe---ThrSerProTyAsnAspAsnTrpAsp 394
 Db 57581 GACGGCAGCGGTACACGCTCAAGTGTACGGCAGCGCGGCGGAGTTCGGAC----- 57631
 QY 395 GlyArgAsnAsnValGluAsnValPhe---IleAsnAlaProGln-----SerGly 410
 Db 57632 -----AACATCAACACCGACGTACTCGGTGAACGCTCTCTCGGAGCGGCCAACGCG 57682
 QY 411 ThrTyThrIleGluVal 416
 Db 57683 ACGTGAAACTCGCGGTG 57700

RESULT 14
 US-10-084-846A-2/c
 ; Sequence 2, Application US/10084846A
 ; Publication No. US20040006026A1
 ; GENERAL INFORMATION:
 ; APPLICANT: WEITNAUER, GABRIELE
 ; APPLICANT: MOHLENWEG, AGNES
 ; APPLICANT: TREFFZER, AXEL
 ; APPLICANT: BECHTHOLD, ANDREAS
 ; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
 ; FILE REFERENCE: 1974-005
 ; CURRENT APPLICATION NUMBER: US/10/084,846A
 ; CURRENT FILING DATE: 2003-02-25
 ; PRIOR APPLICATION NUMBER: PCT/EP01/09815
 ; PRIOR FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: DE 101 09 166.4
 ; PRIOR FILING DATE: 2001-02-25
 ; NUMBER OF SEQ ID NOS: 120
 ; SOFTWARE: Patent in Ver. 3.2
 ; SEQ ID NO 2
 ; LENGTH: 59816
 ; TYPE: DNA
 ; ORGANISM: Streptomyces viridochromogenes
 US-10-084-846A-2

Alignment Scores:
 Pred. No.: 1.04e-21 Length: 59816
 Score: 306.50 Matches: 116
 Percent Similarity: 39.20% Conservative: 51
 Best Local Similarity: 27.23% Mismatches: 154
 Query Match: 13.64% Indels: 105
 DB: 15 Gaps: 19

US-09-985-689a-1 (1-434) x US-10-084-846A-2 (1-59816)

QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 3169 GGGCAGGAGTGAGCGGCTACGTACATCGACACCGCGTC-----CGC 3128
 QY 41 SerMethHisGluAlaPheArgGlyIleThrAlaLeuTyR---AlaLeuGlyArgThr 59
 Db 3127 ATCAACCCACAGCGACTTCGGCGCGCGCGCTCTACGCTTACAGCGCATCGACCAACGAC 3068
 QY 60 AsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
 Db 3067 AACACCGCCAGNAGCGCCAGCGCCAGCGGACGACGTCGCGCGGACGCTCGCGGCAAC 3008
 QY 80 GlySerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 99
 Db 3007 GCC-----TACGCGCTCGCCAGAGAGCAAGATCGTAGCGCTCGCGTCTGTAAC 2957
 QY 100 SerGlyGly-----GlyLeuGlyGlyLeuProSerAsn 110
 Db 2956 AACTCCGGCAGGCACACCGCCAGGTCTCGCGGCATCGACTGGGTCTGCCCGGAC 2897
 QY 111 LeuGlnThrLeuPheSerGlnAlaTyRSerAlaGlyAlaArgIleHisThrAsnSerTrp 130

Db 2896 -----GCCGTCAAGCGCGCGTCCGCCAACATGTCTCCCTC 2864
 QY 131 GlyAlaAlaValAsnGlyAlaTyThrThrAspSerArgAsnValAspAspTyRValArg 150
 Db 2863 GCGCGCGCGCGACACGCGCTCGACACGCGGTACGCAAC-----GCCATG 2816
 QY 151 LysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIle 170
 Db 2815 GCCTCCGCGGTCACCTTCGCGCGCGCGGCAACAGTACGACCAACGCTCCACGAGG 2756
 QY 171 SerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeu 188
 Db 2755 TCA---CCCGCAGCGGTACCGAGGCCATCACGCTCGCGCAGCAGCAGCTCGGACGCC 2699
 QY 189 ArgProSerPheGlySerTyRAlaAspAsnIleAsnHisValAlaGlnPheSerSerArg 208
 Db 2698 AAGCGCGGTACTCTCAACTACGGTCCGTCTCTC----- 2666
 QY 209 GlyProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeu 228
 Db 2665 -----GACCTCTTCGCGCGCGGTTCCTCATCACC 2636
 QY 229 SerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyR 248
 Db 2635 TCGGCC-----TGGAACTCAAGCAGCTCGCGCACC 2606
 QY 249 AlaTyMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu 268
 Db 2605 AACACCATCTCGGTACGTGATGGCACCCTCGCAGCTGGCGGCGCGCGCTC--- 2549
 QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeu 285
 Db 2548 -----CACCTCGCGCGCAACCTCGCGCACCCCGTCCAGGTCCGACCGCGCTCAGC 2495
 QY 286 LysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyRProAsnGlyAsnGln 305
 Db 2494 TCGCGCGCCACACCGCGGTCTGTCACCAACCCCGCGCGCTCGCCCAAC----- 2444
 QY 306 GlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyRValAsnGlu----- 323
 Db 2443 -----CGGCTCTGTACGTCTCGCGCGCGGCAC 2417
 QY 324 -----SerSerSerLeuSerThrSerGlnLysAlaThrTyRSerPheThrAla 339
 Db 2416 GACCACCTCGCGCGCGCTTCGAGAACACCGGTACGACTACGATCAACGCAACTCC 2357
 QY 340 ThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThr 359
 Db 2356 ACGGTGAGTCCCGGTGACGCTCTCGCGCTCTCGCGCAACGCGCTCGCGCTCGCC 2297
 QY 360 AlaSerVal-----ThrLeuValAsnAspLeuAspLeuValIleThrAlaPro 375
 Db 2296 GTAGAGGTCCACATCGTCCACACGTATCGCGCACCTCCAGGTCCAGCTGATCGCCCC 2237
 QY 376 AsnGlyThrGlnTyValGlyAsnAspPhe---ThrSerProTyAsnAspAsnTrpAsp 394
 Db 2236 GACGCGCAGGTACACGCTCAAGTCTGACGCGCACCGCGCGCAGTTCGGAC----- 2186
 QY 395 GlyArgAsnAsnValGluAsnValPhe---IleAsnAlaProGln-----SerGly 410
 Db 2185 -----AACATCAACACCGACGTACTCGGTGAACGCTCTCTCGGAGCGCGCAACGCGC 2135
 QY 411 ThrTyThrIleGluVal 416
 Db 2134 ACGTGAACACTCGCGGTG 2117

RESULT 15
 US-09-927-827-29
 ; Sequence 29, Application US/09927827
 ; Publication No. US20030036176A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bower, Stanley G.

Alignment Scores:			
Pred. No.:	6.96e-21	Length:	3743
Score:	283.00	Matches:	127
Percent Similarity:	37.53%	Conservative:	55
Best Local Similarity:	26.19%	Mismatches:	147
Query Match:	12.59%	Indels:	156
DB:	10	Gaps:	24
US-09-995-689A-1 (1-434) x US-09-927-827-29 (1-3743)			
Qy	21	GlyGlnGlyGlnIleValAlaValAlaAAspThrGlyLeu-----AspThrGly	36
Db	1502	GGCAGCGGCACGGTCGTGGCGGTGATTGATACCGGCATCACAGTCATCGGACCTCAAC	1561
Qy	37	ArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeu	56
Db	1562	GCCAAACATCTCTGGGGCTACGACTTCATCAGCATGCGACCAACCGCATGGCAAC	1621
Qy	57	GlyArgThrAsnAsnAlaAsnAsp-----	64
Db	1622	GCCGCTGACAGCAACGCCCGCGACGAAGGCACGTGTAACCGGCCAACGATCGCGGCC	1681
Qy	65	-----ThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGly	78
Db	1682	GGCATTCCCGCGCCAGCTCCAGTGTGCACGGCACCCCATGTGCGCGCACGGTCGGGCA	1741
Qy	79	AsnGlySerThrAsnLysGlyMetAlaProGlnAla-----	90
Db	1742	GTGACCAACAAACACCACCGCGGTAGCGGCACCGCCCTACGGCGCCAAAGTCTGTACCGGTG	1801
Qy	91	-----AsnLeuValPheGlnSerIleMetAsp	99
Db	1802	CGCGTGTCTGGCAAGTGGGTGGTCTCGGTATCCCGACGCCATCGTCTGGGCC	1861
Qy	100	SerGlyGlyLeuGlyLeuProSerAsnLeuGln-----ThrLeuPheSerGln	117
Db	1862	TCCGGCGCACCGTCAGCGGCATCCCGCCAAATCTAACCCGGCGGAGTGTATCAACATG	1921
Qy	118	AlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAla	137
Db	1922	TCGCTCGCGCGCGCGGTAGTCTGTCCACACCATGACAGAAC---GCCATCAACGGTGGC	1978
Qy	138	TyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeu	157
Db	1979	GTGTGCGCGGCACC-----ACGGTGGTG	2002
Qy	158	PheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLys	177
Db	2003	GTCGCGCGCGCAACGATCGCTCCAATGTGTCCGGT---TCGTCGCGCGCAACTGCGCG	2059
Qy	178	AsnAlaIleThrValGlyAlaThrGluAsn-----LeuArgProSerPheGlySerTyr	195
Db	2060	AACGTGATGCGGTGGCGCCGACCACTTCGGCGGGCGGAAGGCACGTATCCAACTTC	2119
Qy	196	AlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArg	215

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 15, 2004, 23:20:18 ; Search time 2368 Seconds
(without alignments)
5473.052 Million cell updates/sec

Title: US-09-985-689a-1
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSGLY.....EVQAYNPVGPQTFLAIWN 434

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US0985689/runat_10032004.112903.19733/app_query.fasta_1.583
-DB=EST -QPMF=fastap -SUFFIX=xt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pc -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0985689@cgn.1.1135 @runat_10032004.112903.19733 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hrc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hrc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	246	10.9	1605	13	BQ622771 CC Contig
2	241	10.7	640	12	BJ359336 BJ359336
3	225.5	10.0	4198	11	AK029048 Mus muscu
4	224.5	10.0	594	12	BJ393752 BJ393752
5	218.5	9.7	532	29	AL492464 T. brucei
6	207.5	9.2	2141	13	BQ142519 Contig6 M
7	200	8.9	508	28	AQ652212 Sheared D
8	195.5	8.9	601	12	BJ387574 BJ387574
9	195.5	8.7	1002	29	CNS06D58
10	195	8.7	771	14	CA320325 UI-M-FW0-
11	191	8.5	633	12	BJ369190 BJ369190
12	190	8.5	718	12	BI750157 F902.1090
13	185.5	8.3	574	29	TA135H10P
14	184.5	8.2	4662	11	BC060627 Mus muscu
15	182.5	8.1	614	9	AJ273402 AJ273402
16	181	8.1	716	28	BZ893395 HL2.0177
17	180	8.0	641	12	BJ393925 BJ393925
18	180	8.0	665	13	BQ770462 UI-M-F10-
19	179.5	8.0	3091	11	BC011275 Mus muscu
20	177.5	7.9	650	9	AJ274038 AJ274038
21	177.5	7.9	681	14	CB690041 CEST-54-B
22	177	7.9	675	14	CF727824 UI-M-HB0-
23	176	7.8	530	29	CNS01080
24	176	7.8	576	14	CD295943 Strp0691.
25	176	7.8	2121	28	BZ424995 100023066
26	172.5	7.7	580	9	AJ273745 AJ273745
27	172.5	7.7	583	9	AJ273947 AJ273947
28	172.5	7.7	593	9	AJ273918 AJ273918
29	172.5	7.7	601	9	AJ273921 AJ273921
30	171.5	7.6	573	14	CA937626 88V42b10.
31	170.5	7.6	601	9	AJ273050 AJ273050
32	170.5	7.6	712	14	CD311344 Strp0691.
33	170.5	7.6	1029	29	CNS071D8
34	170	7.6	895	13	BQ216158 AGENCOURT
35	170	7.6	1572	11	AV107161 Zea mays
36	169.5	7.5	545	13	BU575479 TGESTzv8
37	169.5	7.5	604	9	AJ273185 AJ273185
38	168.5	7.5	601	12	BJ365857 BJ365857
39	168.5	7.5	609	9	AJ272712 AJ272712
40	168.5	7.5	937	12	BG246418 602360428
41	168	7.5	794	14	CF737198 UI-M-HD0-
42	168	7.5	1050	29	CNS0780L
43	167.5	7.5	564	28	AQ651427 Sheared D
44	167.5	7.5	610	9	AJ274218 AJ274218
45	167	7.4	449	12	BJ359939 BJ359939

ALIGNMENTS

RESULT 1
BQ622771
LOCUS CC Contig67
DEFINITION Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus
CDNA, mRNA sequence.
ACCESSION BQ622771
VERSION BQ622771.1 GI:21649940
KEYWORDS EST
SOURCE Conidiobolus coronatus (Delacroixia coronata)
ORGANISM Conidiobolus coronatus
Eukaryota; Fungi; Zygomycota; Zygomycetes; Entomophthorales;
Ancylistaceae; Conidiobolus.
REFERENCE 1 (bases 1 to 1605)

AUTHORS Freimoser, F.M., Screen, S., Hu, G. and St. Leger, R.J.
TITLE EST analysis of genes expressed by the zygomycete pathogen *Conidiobolus coronatus* during optimized secretion of proteins
JOURNAL Unpublished (2002)
COMMENT Contact: Freimoser, F. M.
 Department of Entomology
 University of Maryland
 4112 Plant Sciences Building, College Park, MD 20742, USA
 Tel: 301 405 16 13
 Fax: 301 314 92 90
 Email: ff34@mail.umd.edu.

FEATURES
 source
 1..1605
 /organism="Conidiobolus coronatus"
 /mol_type="mRNA"
 /strain="ARSEF 512"
 /db_xref="taxon:34488"
 /clone_lib="Conidiobolus coronatus ARSEF 512"
 /note="Vector: Unizap; Conidiobolus coronatus was grown in minimal medium supplemented with Manduca sexta cuticle and peptone for 18 hours. A cDNA library was constructed in the unidirectional Lambda vector Unizap."

ORIGIN

Alignment Scores:
 Pred. No.: 1,528-13 Length: 1605
 Score: 246.00 Matches: 85
 Percent Similarity: 43.9% Conservative: 43
 Best Local Similarity: 29.2% Mismatches: 83
 Query Match: 10.9% Indels: 80
 DB: 13 Gaps: 14

US-09-985-689A-1 (1-434) x B0622771 (1-1605)

QY 6 GlylleValLeuAlaAspValalaGlnSerSer-----TyrGlyLeu----- 19
 Db 718 GGTGTGTGAAGGCTATTGCGGTCCAGCAATGCTCTCGGGTCTTGTCTGTTGGT 777
 QY 20 -----TyrGlyGlnGly 23
 Db 778 CAACTGCTAAGCTCGGATCTGCTCTTACTCTTACACCATGCTGTGCTCAAGGT 837
 QY 24 GlnlleValalaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHis 43
 Db 838 CTCACCTGTTTGGTTTAGACTACTGCTCAATGTCAGCCCAATGAC----- 885
 QY 44 GluAlaPheArgGlyLyslleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAla--- 62
 Db 886 -----TTCGGTGGTCGGCCACT-----TGGGTACTAACACTGCTGGT 924
 QY 63 -----AsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
 Db 925 GGTAGCAACACTGATGCTCACGGTCACGGTACTCCTGCTGCTGCTATTGCTGGT-- 981
 QY 80 GlySerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet--- 98
 Db 982 -----ACCACCTATGGTGTGCCAAGAGCTAACATGTTGCCGTTAAGTCTTAGGT 1035
 QY 99 ---AspSerGlyGlyLeuGlyLeuProSerAsnLeuGlnIntrLeuPheSerGln 117
 Db 1036 GATGATGCTCGGATCATCTCTCGAATTATCTCCGGTATTGACTGCGTGTGTTAAG-- 1092
 QY 118 AlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAla 137
 Db 1093 ---CACTCTGCTGCCAAGAAAGTTATCTCTATGAGTTTAGAGGTGGTAAAGACGATGCT 1149
 QY 138 TyrThrTrpAspSerArgAsnValAspAspTyrValArgLysAsnAspMetTrpIleLeu 157
 Db 1150 CTTACACT-----GCTGTTAACACGCTGTACCAAGGAGGTGTGCTGTTGTC 1200
 QY 158 PheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLys 177
 Db 1201 ---GCTGCTGGTACGAAACAGAGATGCTTGTGGATCTCT---CCCGCTTCTGCTCCT 1254

QY 178 AsnAlaIleThrValGlyAlaThrGlu-----AsnLeuArgProSerPheGlySerTyr 195
 Db 1255 TCCGCCATTACCGTTGGTGGCCACTGATGTCATGATAAAAGGCTTCATTCTTACATTC 1314
 QY 196 AlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArg 215
 Db 1315 GGTAGCTGTGTC----- 1326
 QY 216 IleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAla 235
 Db 1327 -----GATATCTTACTCTCTGGTGTCAACATCTCTCCACC----- 1362
 QY 236 ProAspSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSer 255
 Db 1363 -----TGAAGGGATCTAACACTGCCACCAACCATCTCTGCTACCTCT 1407
 QY 256 MetAlaThrProIleValAlaGlyAsnValAla 266
 Db 1408 ATGGCTTGCCCTCACATTGCTGTTAGCTGCT 1440

RESULT 2

LOCUS BJ395336 640 bp mRNA linear EST 08-MAR-2002
 DEFINITION BJ395336 Dictyostelium discoideum cDNA library, SP Dictyostelium
 discoideum cDNA clone dds38b16 5', mRNA sequence.

ACCESSION BJ395336

VERSION BJ395336.1 GI:19306422

KEYWORDS EST.

SOURCE Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE 1 (bases 1 to 640) Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.

AUTHORS

Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.

TITLE Full length cDNA of Dictyostelium discoideum at the slug stage

JOURNAL Unpublished (2002)

COMMENT Contact: Tadasu Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

FEATURES

Location/Qualifiers

1..640

/organism="Dictyostelium discoideum"

/mol_type="mRNA"

/strain="AX4"

/db_xref="taxon:44689"

/clone="dds38b16"

/sex="mat A"

/dev_stage="Slug stage"

/clone_lib="Dictyostelium discoideum cDNA library, SP"

ORIGIN

Alignment Scores:
 Pred. No.: 1,03e-13 Length: 640
 Score: 241.00 Matches: 69
 Percent Similarity: 49.76% Conservative: 36
 Best Local Similarity: 32.70% Mismatches: 72
 Query Match: 10.73% Indels: 34
 DB: 12 Gaps: 9

US-09-985-689A-1 (1-434) x BJ395336 (1-640)

QY 19 LeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg--- 37

Db 22 TTAAAGAGTAAAGGTCAGATATTGATGTTGCTGATCTGGTTTAGATGGTAGCCATTGT 81

QY 38 -----AsnAspSer-----SerMetHisGluAlaPheArgGly 48

Db 82 TTCTTTTCAGATTCAAAGTATCAATCAATCAATTAATCAAGTGAATGAATAATCATAGAAA 141

```

Qy 49 LysileThralaleuTyAlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsnGlyHis 68
Db 142 GTTGTAACT-----TATATTACTACCATGACAAATGAGATTATGTAATGTCAT 192
Qy 69 GlyThrHisValalaglySerValLeuGlyAsnGly-----Ser 81
Db 193 GSTACACATGTTTGTGCTCTGCAGCAGGTACTCCAGAGATTCTTCATGGGCTATTTC 252
Qy 82 ThrAsnIsglyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 101
Db 253 TCATTTAGTGGTCTTGCATCGCAAGATTGCATTTTATCAT-----CTTTTCATCTGA 309
Qy 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrIleuPheSerGlnAlaTySerAla 121
Db 310 AGTTCGTAACCAACACACCCGAGATGATACAGTCAATGTACAAACCATATATGATGCA 369
Qy 122 GlyAlaArgIleHisThrAsnSerTrpGlyAla-----AlaValAsnGlyAla 137
Db 370 GGTGCAAGATGATGCTGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 429
Qy 138 TyThrThrAspSerArgAsnValAspAspTyValArgLys---AsnAspMetThrIle 156
Db 430 TATCCGATGATGCTGCTGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 489
Qy 157 LeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAla 176
Db 490 CTAAGAGCTGCTGTAAT---AAGAGCTATTGTCATCTTTATTAGCTCAAGCAACAGCT 546
Qy 177 LysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer----- 194
Db 547 AAAATGCAATTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 606
Qy 195 -----TyrAlaAspAsnIleAsn 200
Db 607 TTGGAATATTATGATTCTTCAGATAATGCTAAT 639

RESULT 3
AK029048 4198 bp mRNA linear HTC 18-SEP-2003
LOCUS Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched
DEFINITION library, clone:4732484M11 product:membrane-bound transcription
factor protease, site 1, full insert sequence.
ACCESSION AK029048
VERSION AK029048.1 GI:26325017
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus
ORGANISM Mus musculus
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Komo, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL 20499374
MEDLINE 11042159
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohata, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer

```

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the
Functional annotation of a full-length mouse cDNA collection
Nature 409, 585-690 (2001)

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

6 (bases 1 to 4198)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hasegawa, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayateu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohashi, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, I.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akashira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216]
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.

COMMENT

FEATURES
source

Location/Qualifiers
1..4198
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM,DB:4732484M11"
/db_xref="MGI:2391080"
/db_xref="taxon:10090"
/clone="4732484M11"
/tissue_type="skin"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stages="10 days neonate"
417..3575
/note="unnamed protein product; membrane-bound
transcription factor protease, site 1 (MGD|MGI:1927235,
GS|NM_019709, evidence: BLASTN, 99%, match=3782)
putative"
/codon_start=1
/protein_id="BAC36263.1"
/db_xref="GI:26325018"
/translation="MKLVSTWLVLLVLLGKRLGDLGTRALEKAPCPSCSHLTLLK
VEFSTVVEYEVIVAFNGYFPKARNPFISSALSEVENR11PRNPSSDYPSDFE
VIQKEKQAGLLLEDPHNKRVTPQKVPFRSLKPAESPIVPCNETRSQKQSSR
FLKASLSLGSFGFWHATCRHSRLRLRAIPQVACTQADVLWMCQGTGAPRVAVP
TGLSEKPHFNKNRTNWTNRTLDDGLGHGTAVAGVIASRMCGCFAPDAUHFIR
VFTNQSYTSWFLDAFNALIKVDNISIGGDFDFMDFPVDFDKWELTANNVIMVS
AIGNDPLGYTLNPPADQMDVIGGIDFEDNIARFSPRGMTTWELPCGYGRVKPDIV
TYGVRGSGVKGCRALSGTASPVVAGATVLLVSTVOKRELNVNPSVQKQALIASA
RRLQVNMFEQGHGKLLIRAYOILSSVKQASLSPYIDLTECPVMPYCSQPIYVG
GNPIVNTLLNGMGTGTRIVDKPEWRPPLFQNGDNIEVAFSSVSWFNPSSGYLAISI

CDS

1..3575
/note="unnamed protein product; membrane-bound
transcription factor protease, site 1 (MGD|MGI:1927235,
GS|NM_019709, evidence: BLASTN, 99%, match=3782)
putative"
/codon_start=1
/protein_id="BAC36263.1"
/db_xref="GI:26325018"
/translation="MKLVSTWLVLLVLLGKRLGDLGTRALEKAPCPSCSHLTLLK
VEFSTVVEYEVIVAFNGYFPKARNPFISSALSEVENR11PRNPSSDYPSDFE
VIQKEKQAGLLLEDPHNKRVTPQKVPFRSLKPAESPIVPCNETRSQKQSSR
FLKASLSLGSFGFWHATCRHSRLRLRAIPQVACTQADVLWMCQGTGAPRVAVP
TGLSEKPHFNKNRTNWTNRTLDDGLGHGTAVAGVIASRMCGCFAPDAUHFIR
VFTNQSYTSWFLDAFNALIKVDNISIGGDFDFMDFPVDFDKWELTANNVIMVS
AIGNDPLGYTLNPPADQMDVIGGIDFEDNIARFSPRGMTTWELPCGYGRVKPDIV
TYGVRGSGVKGCRALSGTASPVVAGATVLLVSTVOKRELNVNPSVQKQALIASA
RRLQVNMFEQGHGKLLIRAYOILSSVKQASLSPYIDLTECPVMPYCSQPIYVG
GNPIVNTLLNGMGTGTRIVDKPEWRPPLFQNGDNIEVAFSSVSWFNPSSGYLAISI

SVTKAAWEGIAQOHIMITVASPAETELHSGABHTSTVKLPKIKIPIPTPPSRKVL
 WDOYHNLRYPGYPRDNLRMKNDPLDNGDHVTRFRDMYQHLRSMGYFVEVLGAPF
 TCFDQVGTLLVDSREBEYPPPIAKLRDNDVGLSLVIFSDWNTYSVARKVKEYDE
 NTRDMWEDTGTGANIPALNELLSVNMWGSFGLYEGREVLANDHMYVASGSLAKPEE
 DGVITQTFKQOGLVLEKQETAVENVPIGLYOIPSEGGRIVLGDSNCLDSDHQ
 KDCWLLDALLQYTSYGVTPSLSHSGNRQRPSPAGLAPFERMEGHLHRYISKVLEA
 HLPSPGESGAWDIFPGINPMGRYNOEVGQTPVFVFLGAMVALLAFVVFVQISKASRPKR
 RRPRAKRPQLAQAHAPARTPSV"

polyA_signal

4180..4185

/note="putative"

polyA_site

4198

/note="putative"

ORIGIN

Alignment Scores:

Pred. No.:	8..97e-11	Length:	4198
Score:	225.50	Matches:	116
Percent Similarity:	41.32%	Conservative:	65
Best Local Similarity:	26.48%	Mismatches:	160
Query Match:	10.04%	Indels:	97
DB:	11	Gaps:	25

US-09-985-689A-1 (1-434) x AK029048 (1-4198)

QY 8 ValLysAlaAspValAlaGlnSerSerTyGlyLeuTyGlyGlnGlyGlnLeuValAla 27
 DQ 1005 CTCGAGGAGATGCTGCTGG---CAGATGGATACACAGGTGCTTAATGTCCAGAGTTGCT 1061
 QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
 DQ 1062 GTTTTGTACTGGGCTC-----AGTGAGAAGCATCCGCAATTTTAAG 1103
 QY 48 GlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsn----- 63
 DQ 1104 AAT-----GTGAAGGAGAGAACCAACTGGACCAATGACGGGACCGCTG 1145
 QY 64 ---AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThr 82
 DQ 1146 GATGATGGCTAGGCATCGCACATTCGTGCGGT---GTGATTGCCAGCATGAGGAG 1202
 QY 83 AsnLysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerGly 101
 DQ 1203 TGCCAAAGGATTTGCTCCAGATGACAGGTGCGACATCTTCAGGGTCTTTTACCACAAAT--- 1259
 QY 102 GlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTySerAla 121
 DQ 1260 -----CAGGTGCTTACACATCTGCTTCTGGATGCTTCAACTAT 1301
 QY 122 GlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyThrThrAsp 141
 DQ 1302 GCCATCCTAAAGAAGATGACGCTTCTCAACCTTAGCATCGGTGGGCCGCACTTCATGGAT 1361
 QY 142 SerArgAsnValAspAspTyR-----ValArgLysAsnAspMetThrIleLeuPheAla 159
 DQ 1362 CATCCGTTTGTGACAGGTGGGAATTAACAGCTAACATGTAATGATGTTCTGCT 1421
 QY 160 AlaGlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAla 179
 DQ 1422 ATTGGCAATGAGGACCTCTCATGGCACTCTGAATAACCTCGTCGATCAGATGATGATG 1481
 QY 180 IleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyThrAlaAspAsnIle 199
 DQ 1482 ATTGGAGTGGGTGGGATTGAC-----TTTGAAGATAACATC 1517
 QY 200 AsnHisValAlaGlnPheSerSerArgGly-----ProThrLysAsp 213
 DQ 1518 -----GCTCGCTTTTCTTCCAGGGGAATGACTACTCTGGGAATTTACAGAGGCTAT 1568
 QY 214 GlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSer 233
 DQ 1569 GTCGTGTGAGCCCTGACATTGCTC-----ACCTATGTTGCTGGAGTGGCGGGTTC 1619

QY 234 LeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyAlaTyMetGlyGly 253
 DQ 1620 GGTGTGAAGGGGGGTGC-----CGTGCACTCTCAGGG 1652
 QY 254 ThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheVal 273
 DQ 1653 ACAGGTGTCGTTCCCGAGTGTGCTGGGGCCGTCACCTTGTAGTAGACAGTACAG 1712
 QY 274 LysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla 293
 DQ 1713 AAGGGGAGCTGGTG-----AATCTGCCAGTGTGAAGCAAGCTTTGATACGCTCAGGC 1766
 QY 294 AlaAspIleGlyLeuGlyTyTrpAsnGlyAsn-----GlnGlyTrpGlyArgVal 310
 DQ 1767 CGGAGACTT-----CCTGGGTCAACATGTTCCGACCAAGGTTCATGCCAAGTGTG 1814
 QY 311 ThrLeuAspLysSerLeuAsnVal---AlaTyValAsnGluSerSerSerLeuSerThr 329
 DQ 1815 GATCTGCTCGAGCTATCAGATCTCAGACAGCTATAAACCAGCAGCAAGCTGAGTCT 1874
 QY 330 Ser-----GlnLysAlaThrTySerPheThrAlaThrAlaGlyLysProLeu 345
 DQ 1875 AGCTACATCAGCTGAGTGTGCTCCTACATGTGG---CCCTACTGTCTCCAGCT--- 1928
 QY 346 LysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVal 365
 DQ 1929 -----ATCTACTATGAGGAATGCCA-----ACAATCGTTAATGTCAACCTCCTC 1973
 QY 366 AsnAspLeuAspLeu-----ValIleThrAlaProAsnGlyThrGlnTyValGly 382
 DQ 1974 AATGGCATGGGCTCAGAGGAAGATTTGGTAAAGCTGAGTGGCGACCCCTATTATA--- 2030
 QY 383 AsnAspPheThrSerProTyTrpAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnVal 402
 DQ 2031 -----CCACAGATGGA-----GACACATTCGAGTGGCC 2060
 QY 403 Phe-----IleAsnAlaProGlnSerGlyThrTyThrIleGluVal 416
 DQ 2061 TTCTCCTACTCCTCAGTGTGTGGCCCTGTCAGGTACCTTGCCTCCTCCATT 2114

RESULT 4

BJ393752

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..594

/organism="Dictyostelium discoideum"

/mol_type="rRNA"

/strain="AX4"

/db_xref="taxon:44689"

/clone="dds32b16"

/sex="mat A"

/dev_stage="slug stage"

/clone_lib="Dictyostelium discoideum cDNA library, SF"

ORIGIN

Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approximately 1-4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999).

Email: nleaves@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available at <http://www.sanger.ac.uk/projects/Tbrucei/>.

FEATURES

source

```

1. 532
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TRE927"
/db_xref="taxon:5691"
/clone="319a10"

```

ORIGIN

Alignment Scores:	1.4e-11	Length:	532
Pred. No.:		Matches:	61
Score:	218.50	Conservative:	28
Percent Similarity:	48.3%	Mismatches:	68
Best Local Similarity:	33.1%	Indels:	27
Query Match:	9.72%	Gaps:	6
DB:	29		

US-09-985-689A-1 (1-434) x TA319G10P (1-532)

93	ValPheGlnSerIleMetAspSerGlyGlyGlyLeuGly-----	105
QY		
2	GTAATGCCAAGTATAATCGTTCGCCAGGGGGCAAGATTCTCTCAGGGGTGGCGTGCC	61
Db		
106	-----GlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyr	119
QY		
62	CATCCCACTACAGCTTGCTCTCCCCACGACGTACTCAATATTTCGTCCCGGTATAT	121
Db		
120	SerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr	139
QY		
122	GCGCGTGGAGCCCGTGTCTCTCAAACTCGTGGGGTTTGTCTCCCTCCGAGTATTCT	181
Db		
140	ThrAspSerArgAsnValAspAspTyrValArgLys---AsnAspMetThrIleLeuPhe	158
QY		
182	GCTGTGGAAAAGGATATCGATAGTATTCGCGAGTAGTTATGACGATCGCTATCTTC	241
Db		
159	AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsn	178
QY		
242	TCCACTGGCAACAGTATCCCAAGATGGC-----CTAATGACTCCGTGCTGGTAGAAC	295
Db		
179	AlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsn	198
QY		
296	GTGATGTGCGTGGGTGCACACAAAACGTG-----TTTGACGCTTCGAAAGAC---	343
Db		
199	IleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysPro	218
QY		
344	-----ATTGTTTCTTCGTTTCTTCGATGGTCCCAACATACGACGTTAGGATGAACCC	397
Db		
219	AspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSer	238
QY		
398	GATCTTGTCGTCCTCCGGGAAGAGTGGTGTCTGCTCTTCTCTCGCAAGCATCAGCT	457
Db		
239	SerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThr	258
QY		
458	AAA-----CAATGTAAAGTGGTGCCCAAGCGGNGTTTCATCGATGGCACT	502
Db		
259	ProfileValAla 262	
QY		
503	GCGCGCGTCGGG 514	
Db		

RESULT 6
BQ142519/c
LOCUS

EST 24-APR-2002

mRNA

linear

EST 24-APR-2002

Saccharomycetales; Saccharomycetaceae; Zygosaccharomyces.

REFERENCE
AUTHORS

1 (Dases 1 to 1002)

1. *Chromomyces*, *Saccaria Mycetaceae*, *Zygomycota*, *Myxomycetes*.

Sources: J.L., Agle, M., Artigueave, F., Blandin, G., Boletín-Fukuhara, M., Bon, E., Brothier, P., Casaregola, S., De-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B., Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekai, F., Toffian-Nioche, C., Wesolowski-Louvel, M., Wincker, P. and Weissenbach, J.

TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)

FEB 5 1968

p. 487 (1), 3-12 (2000)

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

MEDLINE
PUBMED

20584711
11152876
2 (bases 1 to 1002)
de Montigny, J., Straub, M., Potier, S., Tekala, F., Dujon, B.,
Wincker, P., Artiguenave, F. and Souciet, J.
Genomic exploration of the hemiascomycetous yeasts: 8.

Zygobacc

pharymyces rouxi

recombinant γ

0. 00000

JOURNAL
MEDLINE

FEBS Lett. 487 (1), 52-55 (2000)

t. 487 (1), 52-55 (2000)

MEDLINE
PUBMED

20584718
11152883

3 (bases 1 to 1002)
Genoscope.
Direct Submission
Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremlieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segf@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces*
exiguus, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*,
Saccharomyces kluyveri, *Kluyveromyces thermotolerans*, *Kluyveromyces*
lactis var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia*
angusta, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*,
Candida tropicalis and *Yarrowia lipolytica*. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES

```

location/Qualifiers
1..1002
/organism="Zygosaccharomyces rouxii"
/mol_type="genomic DNA"
/strain="CBS 732"
/db_xref="taxon:4956"
/clone="AR0AA010H04"
/clone_lib="AR0AA"
/note="end : T3"
<2..>979
/note="similar to Saccharomyces cerevisiae ORF YCR045c [
similarity to serin proteases ]"
/evidence="not experimental"

```

ORIGIN

```

/revdence=not_experimental

scores:
      8.49e-09      Length:      1002
      195.50      Matches:      81
      41.90%      Conservative: 38
      28.52%      Mismatches: 102
      8.70%      Indels:      63
      29      Gaps:      14

39A-1 (1-434) x CNS06D6B (1-1002)

```

QY	14	GlnSerSer	ThrGlyLeu	TYR-----	GlyGlnGlyGlnIleVal	Ala	27
		:::	:::	:::	:::		
Db	278	GAAGAGAG	TTTCAATT	TATTACTATT	TAAATG	GCACAA	GGGTAGAAATATCAATGCATAT
		:::	:::	:::	:::		
QY	28	ValAlaAsp	ThrGlyLeu	AspThrGlyArg	AsnAspSerSer	MetHisGlnAla	pheArg
		:::	:::	:::	:::	:::	
Db	338	ATWATG	TATCTG	GAATC-----	TATAAG	ACACAT	AAGCACTTCGT
		:::	:::	:::	:::		
QY	48	GlyIysIle	ThrAlaLeu	TYrAlaLeu	GlyArgThr	AsnAsnAla---	AsnAspThrAsn
		:::	:::	:::	:::		
Db	380	GGACGT-----	GCATATTCGG	CGGAGAGACTTT	ACAGGCG	GAAGGCTCTTG	GTGATCGTAAT
		:::	:::	:::	:::		

http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pYX-5.

FEATURES

Location/Qualifiers

```
1. .771
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6816072"
/tissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="PH108 (T1 phage resistant)"
/clone_lib="NIH BMAP FWO"
/Note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCGAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."
```

ORIGIN

Alignment Scores:
Pred. No.: 6,14e-09 Length: 771
Score: 195.00 Matches: 82
Percent Similarity: 41.81% Conservative: 38
Best Local Similarity: 28.57% Mismatches: 109
Query Match: 8.68% Indels: 58
DB: 14 Gaps: 13

US-09-985-689A-1 (1-434) x CA320325 (1-771)

```
QY 8 ValLysAlaAspValAlaGlnSerSerTyGlyLeuTyGlyGlnGlyValAla 27
Db 40 CTGAGGAGAGATGCTGTGG---CAGATGGGATACACAGGTGCTAATGTCAGAGTTGCT 96
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 97 GTTTTGTACTGGGCTC-----AGTGAGAAGCATCCGCAATTTTAAAG 138
QY 48 GlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsn----- 63
Db 139 AAT-----GTGAGGAGAGAGACCACTGGACCACTGAGCGGACCCCTG 180
QY 64 ---AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThr 82
Db 181 GATGATGGGCTAGGCCATGCCACATTCGTTGCAGGT---GTGATGCCAGCATGAGGGAG 237
QY 83 AsnLysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerGly 101
Db 238 TGCACAGGATTGCTCCAGATGCAGAGCTGCACATCTTCAGGTCTTTACCAACAAAT--- 294
QY 102 GlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTySerAla 121
Db 295 -----CAGGTGTCTTACACATCTTGGTTTCTGGATGCCCTCAACTAT 336
QY 122 GlyAlaArgIleHisThrAsnSerTPGlyAlaAlaValAsnGlyAlaTyThrThrAsp 141
Db 337 GCATCTCTAAGAAGATGGACGTTCTCAACCTTAGCATCGGTGGGCCCGCACTTCATGGAT 396
QY 142 SerArgAsnValAspAspTyTyr-----ValArgLysAsnAspMetThrIleLeuPheAla 159
Db 397 CATCCGTTTGTGCACAGGTGGGAATTAAACAGCTAAACAATGTAATTATGTTTCTGCT 456
```

```
QY 160 AlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla 179
Db 457 ATTGGCAATGATGACCTCTCTATGGCACTCTCTGAATACCTGCTGATCAGATGGATGTG 516
QY 180 IleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIle 199
Db 517 ATTGGAGTGGGTGGCATTTGAC-----TTTGAAGATAACATC 552
QY 200 AsnHisValAlaGlnPheSerSerArgGly-----ProThrLysAsp 213
Db 553 -----GCTCGCTTTTCTCCAGGGGAATACCTCTGGGAATTACACAGAGGGGTAT 603
QY 214 GlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSer 233
Db 604 GGTGCTGTGAGCCTGACATTGTC-----ACCTATGCTGCTGGAGTCCGGGGTTC 654
QY 234 LeuAlaProAspSerSerPheTPAlaAsnHisAspSerLysTyAlaTyMetGlyGly 253
Db 655 GGTGTGAAAGGGGGCTGC-----CGTGCACTCTCAGGG 687
QY 254 ThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheVal 273
Db 688 ACCAGTGTCTTCCCTCCAGTGTGCTGGGGGGTGCACCTTTAGTANGCAGACAGTACAG 747
QY 274 LysAsnArgGlyIleThrPro 280
Db 748 AAGCGGAGCTGTGTAATCCT 768

RESULT 11
BJ369190
LOCUS BJ369190
DEFINITION BJ369190 Dictyostelium discoideum cDNA library, CF Dictyostelium
ACCESSION BJ369190
VERSION BJ369190.1 GI:19278573
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 633)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the culmination
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
source
location/Qualifiers
1. .633
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc49116"
/sex="mat A"
/dev_stage="Culmination stage"
/clone_lib="Dictyostelium discoideum cDNA library, CF"

ORIGIN
Alignment Scores:
Pred. No.: 1,11e-08 Length: 633
Score: 191.00 Matches: 57
Percent Similarity: 41.55% Conservative: 29
Best Local Similarity: 27.54% Mismatches: 49
Query Match: 8.50% Indels: 72
DB: 12 Gaps: 6

US-09-985-689A-1 (1-434) x BJ369190 (1-633)
```


Fishler, K., Keppel, C., Kucaba, T., Lebeck, M., Melo, A., Schaefer, K.,
Scheetz, T., Smith, C., Snir, E., Tack, D., Trout, K., Walters, J.,
Casavant, T., Soares, M.B.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: Plate: Row: Column: 0
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 6678418
This clone has the following problem: frame shifted.

FEATURES

source

Location/Qualifiers
1..4662
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6844459"
/tissue_type="Brain, mouse, 13.5,14.5,16.5,17.5 dpc"
/clone_lib="NIH SWAP_FYO"
/lab_host="DH10B"
/note="Vector: pYX-ASC"

ORIGIN

Alignment Scores:
Pred. No.: 1.45e-06 Length: 4662
Score: 184.50 Matches: 101
Percent Similarity: 35.56% Conservative: 69
Best Local Similarity: 21.13% Mismatches: 155
Query Match: 8.21% Indels: 153
Gaps: 21

US-09-985-689A-1 (1-434) x BC060627 (1-4662)

QY 10 AlaAspValAlaGlnSerSerTyrglyLeuTyrglyGlnGlyGlnIleValAlaValAla 29
|||||
Db 803 GCTGAGATGCTCAATTACTCTGTGAACATTATGACATGGGAACCTGCTCCATTGTG 862
QY 30 AspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLys 49
|||||
Db 863 ACCAGCGGA----- 871
QY 50 IleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsnGlyHisGly 69
|||||
Db 872 -----GGAGCTCATGGA 883
QY 70 ThrHisValAlaGlySerValLeuGlyAsn-----GlySerThrAsnLysGlyMet 86
|||||
Db 884 ACCCATGTAGCAAGTATAGCCGACGGCATTTTCCAGAGAGCCCTGAACCGAATGGAGTT 943
QY 87 AlaProGlnAlaAsnLeuValPhe-----GlnSerIleMetAsp 99
|||||
Db 944 GCTCTGTGTCTCAATTCTATCCATTAAAGATTGGTGATACACGGCTAAGCACTATGGAA 1003
QY 100 SerGlyGlyLeuGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAla--- 118
|||||
Db 1004 ACAGGCACAGGCGCTC-----ATCAGAGCTATGATAGAAGTTATAAATCATCAAGTGTGAT 1057
QY 119 -----TySerAlaGlyAlaArgIleHisThrAsnSerTipGlyAlaAlaValAsn 135
|||||
Db 1058 CTTGTCACTACAGTTATGGAGAACCACTATTGGCCCAANTTTGGAGAAATT----- 1111
QY 136 GlyAlaTyThrThrAspSerArgAsnValAspAspTyValArgLysAsnAspMetThr 155
|||||
Db 1112 -----TGTAAGTAATTAATGAAGCAGTATGGAAACATAATACAAATT 1153
QY 156 IleLeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaPro---Gly 174
|||||
Db 1154 TATGTTTCAAGCGCTGGAAATTAATGTCATGCCCTTTCTACAGTGGGTGTCCAGAGAGA 1213
QY 175 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer 194
|||||
Db 1214 ACTACATCCAGTGTATAGTGTGGAGCT----- 1243

QY 195 TYRAlaAspAsnIleAsnHisValAlaGlnPhe----- 205
|||||
Db 1244 TATGTTTCCCTGATATGATGTTCCAGAGATTTCACATGAGAGAAACTCCCTGGCAAT 1303
QY 206 -----SerSerArgGlyProThrLysAspGlyArgIleLysProAspValMet 221
|||||
Db 1304 CAATATACATGCTCTCTAGAGGCCCAAGTGTATGAGGCCCTCGGTGTGAGCATAGT 1363
QY 222 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 241
|||||
Db 1364 GCACCAGGAGGTGATTGTTCTGTGCTAAC-----TGG 1399
QY 242 AlaAsnHisAspSerLysTyAlaTyMetGlyGlyThrSerMetAlaThrProIleVal 261
|||||
Db 1400 ACATTGAGGGGACTCAG-----CTAATGAATGGACATCAATGCTCTCCCAATGCC 1453
QY 262 AlaGlyAsnValAla-----GlnLeuArgGluHisPheValLysAsnArgGly 277
|||||
Db 1454 TGTGTGGCATTCCTCGTACTTTTCAGGGCTGAAGCAATAATATTTGACTATCTACTGTA 1513
QY 278 IleThrProLysProSerLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGly 297
|||||
Db 1514 CACTCAGTCAGAGAGCTCTAGAAAATACTCAATATAAAGCTGACAAATATAGAACTATT 1573
QY 298 LeuGlyTyThrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsn 317
|||||
Db 1574 GCC-----CAAGGACATGGAATATTTCAGGTGTGACAAAGCTTATGAC 1615
QY 318 ValAlaTyValAsnGluSerSerSerLeuSerGlnLysAlaThrTyThrSerPhe 337
|||||
Db 1616 TACCTCAATCAAAAT-----ACATCATTTGCTAACAGATTAGTTT 1657
QY 338 ThrAlaThrAlaGlyLys-----ProLeuValIleSerLeu 349
|||||
Db 1658 ACAGTTACTGTTGGAAATAACCGTGTATCTACCTCCGAGATCCCTGTCCAGTGGCTGCT 1717
QY 350 ValTrpSerAsp-----AlaProAlaSerThrThrAlaSer 361
|||||
Db 1718 CCT---TCAGATCATGGTGTGGCATTGCGCTGATTTCAGAGAACACAGAAACTCT 1774
QY 362 ValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyVal 381
|||||
Db 1775 GAAAAATATCTTTTCAGCTTCATTAGCTTTAACTTCA----- 1813
QY 382 GlyAsnAspPheThrSerProTyThrAsnAspAsnTrp-----Asp 394
|||||
Db 1814 -----AATTCATCTTGGTTCAGTGTCCAGGCCATTGGAA 1849
QY 395 GlyArgAsnAsnValGluAsnValPheIleAsnAlaPro-----GlnSerGly 410
|||||
Db 1850 CTCATGAATCAGTGTCCGACATAAACATACGTGTGGACCCCGAGGGGCTTAAGAGAGGG 1909
QY 411 ThrTyThrIleGluValGlnAlaTyThrAsnValPro-----ValGlyPro 425
|||||
Db 1910 TTACATTATACAGAGTATGTGGCTATGATATAGCATCCCAATGCAAGTCT 1963

RESULT 15

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AJ273402 614 bp mRNA linear EST 29-DEC-1999
AJ273402 Metarhizium anisopliae ARSEF 2575 Metarhizium anisopliae
CDNA clone M4948, mRNA sequence.

AJ273402
AJ273402.1 GI:6432774

EST.

Metarhizium anisopliae

Metarhizium anisopliae

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

Hypocreomycetidae; Hypocreales; Clavicipitaceae; Mitosporic

Clavicipitaceae; Metarhizium.

1 (bases 1 to 614)

Screen, S.E., Mathur, P. and St. Leger, R.J.

EST analysis of the insect pathogenic fungus Metarhizium anisopliae

Unpublished (1999)

COMMENT Contact: Screen SE

Entomology
University of Maryland
4112 Plant Sciences Building, College Park, MD 20742, USA.
Location/Qualifiers

FEATURES

source
1..614
/organism="Metarhizium anisopliae"
/mol_type="mRNA"
/strain="ARSEP 2575"
/db_xref="taxon:5530"
/clone="Ma#948"
/clone_lib="Metarhizium anisopliae ARSEP 2575"
/note="Vector: Unizap; Metarhizium anisopliae was grown on insect cuticle for 24 hours. A cDNA library was constructed in the unidirectional Lambda vector, Unizap"

ORIGIN

Alignment Scores:

Pred. No.:	7.61e-08	Length:	614
Scores:	182.50	Matches:	67
Percent Similarity:	42.86%	Conservative:	38
Best Local Similarity:	27.24%	Mismatches:	88
Query Match:	8.12%	Indels:	53
DB:	9	Gaps:	12

US-09-985-689A-1 (1-434) x AJ273402 (1-614)

```
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 11 GGTGAGGTACTTCGGTATATATATCATTCGACACTGGTATTGAG-----52
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 53 GCCTCCACCCCGAGTTGAGGTCGGCCACITTTCTTAAGAGCTTCATCAGCGGTCAA 112
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 113 AAC---ACTGATGCCACGGCCATGGGACTCACTGCGCTGTACCAATT-----GGT 160
Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 161 AGCAAGACCTACGGTGTGGCCAAAAGCTAAGCTCTATGGTGTCAAGGTTCTTGACAAC 220
Qy 101 -----GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAla 118
Db 221 CAGGCGCAGTGTTCTCTACTCCGTATCATCAGTGGCATGGACTAGCTTGCACAGGACTCC 280
Qy 119 TyrSerAlaGly-----AlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsn 135
Db 281 AAGACCGCGGCTCCCAACGGCGCCATTCCTTCATGAGCCTGGGA-----328
Qy 136 GlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThr 155
Db 329 GGTGGTACTCGGGTCCGTCACCAAGGTGCTGCTGTTGGTCAATTCTGGTGTCTTC 388
Qy 156 IleLeuPheAlaAlaGlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGlyThr 175
Db 389 CTTCCGCTGCGGCTGGCAACGATAACCGGATGCCCAACACCTCT---CCCGCTTCC 445
Qy 176 AlalysAsnAlaIleThrValGlyValAlaThr-----GluAsnLeuArgProSerPheGly 193
Db 446 GAGCCTTCTGCTGCATCTGTGGTCTCTGCGGAAAATGACAGCGCATCTTCTCTCTTC 505
Qy 194 SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAsp 213
Db 506 AACTAC-----511
Qy 214 GlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSer 233
Db 512 GGCAGAGTT---GTCGATATTTTCGTCCTCTGGTAGCAATGTTCTTTCACC-----559
Qy 234 LeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGly 253
Db 253
```

Db 560 -----TGGATTGGT-----GGCCGCACAAACACCATCTCTGGT 592

Qy 254 ThrSerMetAlaThrPro 259

Db 593 ACCTNCAITGGCTACTCCC 610

Search completed: March 16, 2004, 01:15:53
Job time : 2379 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:46:53 ; Search time 59 Seconds
(without alignments)
2078.400 Million cell updates/sec

Title: US-09-985-689a-1-COPY
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSVGLY.....EVQAYNPVCPQTFSLAIVN 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB ID	Description
1	2242	99.8	434	5	AAW50080	Bacillus
2	2242	99.8	640	2	AAV17090	Bacillus
3	2237	99.5	640	2	AAV17091	Bacillus
4	2186	97.3	434	5	AAW50081	Bacillus
5	2178	96.9	639	2	AAV17089	Bacillus
6	2150	95.7	639	2	AAV17087	Bacillus
7	2150	95.7	640	2	AAV17088	Bacillus
8	2138	95.1	434	5	AAW50085	Bacillus
9	2120.5	94.4	433	5	AAW50086	Bacillus
10	2120.5	94.4	641	2	AAW89547	Bacillus
11	2071	92.2	434	5	AAW50090	Bacillus
12	1933.5	88.7	433	5	AAW50084	Bacillus
13	1989.5	88.5	433	5	AAW50082	Bacillus
14	1984.5	88.3	636	2	AAW89548	Bacillus
15	1982.5	88.2	433	5	AAW50083	Bacillus
16	1981.5	88.2	433	2	AAW26274	Alkali-pr
17	1981.5	88.2	433	2	AAW61495	Modified
18	1981.5	88.2	433	2	AAW95698	Bacillus
19	1981.5	88.2	433	3	AAV69207	Amino aci
20	1981.5	88.2	433	3	AAV44619	Bacillus
21	1581.5	70.4	345	2	AAW62230	Subtilase
22	1581.5	70.4	345	2	AAW21654	Subtilase
23	447.5	19.9	659	2	AAW24121	Thermococ
24	447.5	19.9	659	2	AAW94840	WO9856926
25	409	16.2	412	2	AAW94836	Hyperther

26	409	18.2	522	2	AAW24122	Pyrococcus
27	409	18.2	522	2	AAW94838	Hyperther
28	409	18.2	654	2	AAW24129	Pyrococcus
29	409	18.2	654	2	AAW94841	Hyperther
30	396	17.6	659	2	AAW24123	Protease.
31	372	16.6	545	4	ABW09483	T. yonsei
32	366	16.3	1079	4	ABW09483	Transglut
33	366	16.3	1079	6	ABW09483	Transglut
34	348	15.5	734	2	AAW13666	Foreign p
35	348	15.5	734	2	AAW13666	Fragment.
36	348	15.5	823	2	AAW13668	Streptomy
37	336	15.0	1237	6	ABU11343	Dhpa-mel
38	311.5	13.9	806	2	AAW27481	Protein e
39	306.5	13.6	519	6	ABP76735	RP-III re
40	306.5	13.6	19938	6	ABP76735	Streptomy
41	305.5	13.6	903	2	AAW87007	Streptomy
42	305.5	13.6	1398	2	AAW87008	Hyperther
43	305.5	13.6	1398	2	AAW24124	Protease.
44	305.5	13.6	1398	2	AAW94839	Pyrococcus
45	296.5	13.2	699	2	AAW08471	WO9856926
					AAW08471	F. balust

ALIGNMENTS

RESULT 1
AAW50080
ID AAW50080 standard; protein; 434 AA.
XX
AC AAW50080;
XX
DT 12-AUG-2002 (first entry)
XX
DE Bacillus sp KSM-KP43 alkaline protease protein fragment.
XX
KW Alkaline protease; detergent; laundry; bleaching; dishwasher.
XX
OS Bacillus sp.
XX
PN EP1209233-A2.
XX
PD 29-MAY-2002.
XX
PF 22-NOV-2001; 2001EP-00127851.
XX
PR 22-NOV-2000; 2000JP-00355166.
PR 12-APR-2001; 2001JP-00114048.
XX
PA (KAOS) KAO CORP.
XX
PI Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
PI Okuda M, Saeki K;
XX
DR WPI; 2002-437518/47.
XX
PT New modified alkaline proteases useful in detergent compositions.
XX
PS Claim 1; Page 10-11; 25pp; English.
XX
CC This invention describes novel Bacillus sp. alkaline proteases useful in
CC detergent compositions, especially in laundry, bleaching or automatic
CC dishwasher detergents. The novel proteases have an increased detergency %
CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
CC sequence represents a fragment of the alkaline protease KP43 from
CC Bacillus sp strain KSM-KP43 which is used to create the modified protease
CC represented in AAW50090
XX
SQ Sequence 434 AA;

Query Match 99.8%; Score 2242; DB 5; Length 434;
Best Local Similarity 99.8%; Pred. No. 7.2e-152;
Matches 433; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
 Db 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
 QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
 Db 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
 QY 121 AGARIHTNSWGAANGAYTTDSRNVDVYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
 Db 121 AGARIHTNSWGAANGAYTTDSRNVDVYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
 QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 240
 Db 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 240
 QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 Db 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 QY 301 PNGNOGWGRVTLDKSLNVAAYVNESSLSSTQKATYSFTATAGKPLKISLVMSDAPASTTA 360
 Db 301 PNGNOGWGRVTLDKSLNVAAYVNESSLSSTQKATYSFTATAGKPLKISLVMSDAPASTTA 360
 QY 361 SVTLVNDLNLVITAPNGTQVGNDFTPSYNDNDGNNVENVFINAPOSGTYYTIEVQAYN 420
 Db 361 SVTLVNDLNLVITAPNGTQVGNDFTPSYNDNDGNNVENVFINAPOSGTYYTIEVQAYN 420
 QY 421 VPVGPOTFSLAIVN 434
 Db 421 VPVGPOTFSLAIVN 434

RESULT 2

AAV17090
 ID AAY17090 standard; protein; 640 AA.
 AC AAY17090;
 XX
 DT 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 DE Bacillus alkaline protease.
 XX Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent.
 XX
 OS Bacillus sp.
 XX
 PN WO9918218-A1.
 XX
 PD 15-APR-1999.
 XX
 PF 07-OCT-1998; 98WO-JP004528.
 XX
 PR 07-OCT-1997; 97JP-00274570.
 XX
 PA (KAOS) KAO CORP.
 XX
 PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX
 DR WPI; 1999-287736/27.
 DR N-PSDB; AAX37278.
 XX
 PT Alkali protease from Bacillus used in washing powders.
 XX
 PS Disclosure; Page 58-63; 71pp; Japanese.
 XX

CC The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it

CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease. (Updated
 CC on 20-MAR-2003 to correct DR field.)
 XX

SO Sequence 640 AA;

Query Match 99.8%; Score 2242; DB 2; Length 640;
 Best Local Similarity 99.8%; Pred. No. 1.2e-151;
 Matches 431; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
 Db 207 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 266
 QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
 Db 267 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 326
 QY 121 AGARIHTNSWGAANGAYTTDSRNVDVYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
 Db 327 AGARIHTNSWGAANGAYTTDSRNVDVYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 386
 QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 240
 Db 387 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 446
 QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 Db 447 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 506
 QY 301 PNGNOGWGRVTLDKSLNVAAYVNESSLSSTQKATYSFTATAGKPLKISLVMSDAPASTTA 360
 Db 507 PNGNOGWGRVTLDKSLNVAAYVNESSLSSTQKATYSFTATAGKPLKISLVMSDAPASTTA 566
 QY 361 SVTLVNDLNLVITAPNGTQVGNDFTPSYNDNDGNNVENVFINAPOSGTYYTIEVQAYN 420
 Db 567 SVTLVNDLNLVITAPNGTQVGNDFTPSYNDNDGNNVENVFINAPOSGTYYTIEVQAYN 626
 QY 421 VPVGPOTFSLAIVN 434
 Db 627 VPVGPOTFSLAIVN 640

RESULT 3

AAV17091
 ID AAY17091 standard; protein; 640 AA.
 AC AAY17091;
 XX
 DT 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 DE Bacillus alkaline protease.
 XX
 KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent.
 XX
 OS Bacillus sp.
 XX
 PN WO9918218-A1.
 XX
 PD 15-APR-1999.
 XX
 PF 07-OCT-1998; 98WO-JP004528.
 XX
 PR 07-OCT-1997; 97JP-00274570.

XX PA (KAOS) KAO CORP.
XX PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
XX PI Shikata S, Nomura M;
XX DR WPI; 1999-287736/27.
XX DR N-PSDB; AAX37279.
XX PT Alkali protease from Bacillus used in washing powders.
XX PS Disclosure; Page 63-68; 7lpp; Japanese.
XX CC The invention relates to alkaline proteases produced by strains of
CC Bacillus. The proteases ability to digest casein is not inhibited by
CC oleic acid and they have a high stability to oxidizing agents. The
CC alkaline protease of the invention has the following properties: (a) it
CC is active over the pH range 4-13 and has at least 80% of its optimum
CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
CC its ability to digest casein is not inhibited by oleic acid; (e) it has
CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
CC used as enzymes in washing compositions for use in automatic dishwashers
CC and for washing clothes. The stability to oxidising agents allows the
CC enzyme to be an effective component of washing compositions including
CC bleaches. The present sequence represents an alkaline protease. (Updated
CC on 20-MAR-2003 to correct DR field.)
XX SQ Sequence 640 AA;

Query Match 99.6%; Score 2237; DB 2; Length 640;
Best Local Similarity 99.3%; Pred. No. 2,7e-151;
Matches 432; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 NDVARGIVKADVAQSSYGLYGQGI VAVADTGLDGRNDSSMHEAFRGKITALYALGRN 60
DB 207 NDVARGIVKADVAQSSYGLYGQGI VAVADTGLDGRNDSSMHEAFRGKITALYALGRN 266
QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 267 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 326
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
DB 327 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 386
QY 181 TVGATENLRPSFGSYADNINHVAFSGRGPTKDGRIKPDVMAPTTILSARSLAPDSSF 240
DB 387 TVGATENLRPSFGSYADNINHVAFSGRGPTKDGRIKPDVMAPTTILSARSLAPDSSF 446
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
DB 447 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 506
QY 301 PNGNQGWGRVTLDKSLNVAIVYNESSLSSTQKATYFTATAGKPLKISLWSDAPASTTA 360
DB 507 PNGNQGWGRVTLDKSLNVAIVYNESSLSSTQKATYFTATAGKPLKISLWSDAPASTTA 566
QY 361 SVTLVNDLNLVITAPNGTQVVGNDFTSPYNDNDGRNNVENVFINAPQSGTYTIEVQAYN 420
DB 567 SVTLVNDLNLVITAPNGTQVVGNDFTSPYNDNDGRNNVENVFINAPQSGTYTIEVQAYN 626
QY 421 VPVGQPTFSLAIVN 434
DB 627 VPVGQPTFSLAIVN 640

RESULT 4
AAM50081
ID AAM50081 standard; protein; 434 AA.
XX
AC
XX AAM50081;

DT 12-AUG-2002 (first entry)
XX DE Bacillus sp KSM-KP9860 alkaline protease protein fragment.
XX KW Alkaline protease; detergent; laundry; bleaching; dishwasher.
XX OS Bacillus sp.
XX PN EP1209233-A2.
XX PD 29-MAY-2002.
XX PF 22-NOV-2001; 2001EP-00127851.
XX PR 22-NOV-2000; 2000JP-00355166.
XX PR 12-APR-2001; 2001JP-00114048.
XX PA (KAOS) KAO CORP.
XX PI Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
XX PI Okuda M, Saeki K;
XX DR WPI; 2002-437518/47.
XX PT New modified alkaline proteases useful in detergent compositions.
XX PS Claim 5; Page 12-13; 25pp; English.
XX CC This invention describes novel Bacillus sp. alkaline proteases useful in
CC detergent compositions, especially in laundry, bleaching or automatic
CC dishwasher detergents. The novel proteases have an increased detergency
CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
CC sequence represents a fragment of the alkaline protease KP9860 from
CC Bacillus sp strain KSM-KP9860 described in the method of the invention
XX SQ Sequence 434 AA;
Query Match 97.3%; Score 2186; DB 5; Length 434;
Best Local Similarity 96.3%; Pred. No. 7,2e-148;
Matches 418; Conservative 14; Mismatches 2; Indels 0; Gaps 0;
QY 1 NDVARGIVKADVAQSSYGLYGQGI VAVADTGLDGRNDSSMHEAFRGKITALYALGRN 60
DB 1 NDVARGIVKADVAQSSYGLYGQGI VAVADTGLDGRNDSSMHEAFRGKITALYALGRN 60
QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
DB 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
QY 181 TVGATENLRPSFGSYADNINHVAFSGRGPTKDGRIKPDVMAPTTILSARSLAPDSSF 240
DB 181 TVGATENLRPSFGSYADNINHVAFSGRGPTKDGRIKPDVMAPTTILSARSLAPDSSF 240
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
DB 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
QY 301 PNGNQGWGRVTLDKSLNVAIVYNESSLSSTQKATYFTATAGKPLKISLWSDAPASTTA 360
DB 301 PNGNQGWGRVTLDKSLNVAIVYNESSLSSTQKATYFTATAGKPLKISLWSDAPASTTA 360
QY 361 SVTLVNDLNLVITAPNGTQVVGNDFTSPYNDNDGRNNVENVFINAPQSGTYTIEVQAYN 420
DB 361 SVTLVNDLNLVITAPNGTQVVGNDFTSPYNDNDGRNNVENVFINAPQSGTYTIEVQAYN 420
QY 421 VPVGQPTFSLAIVN 434
DB 421 VPVGQPTFSLAIVN 434

RESULT 5
 AAY17089
 ID AAY17089 standard; protein; 639 AA.
 XX
 AC AAY17089;
 XX
 DT 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 DE Bacillus alkaline protease.
 XX
 KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent.
 XX
 OS Bacillus sp.
 XX
 PN WO9918218-A1.
 XX
 PD 15-APR-1999.
 XX
 PF 07-OCT-1998; 98WO-JP004528.
 XX
 PR 07-OCT-1997; 97JP-00274570.
 XX
 PA (KAOS) KAO CORP.
 XX
 PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX
 DR WPI; 1999-287736/27.
 DR N-PSDB; AAX37277.
 XX
 PT Alkali protease from Bacillus used in washing powders.
 XX
 PS Disclosure; Page 53-58; 71pp; Japanese.
 XX
 CC The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease. (Updated
 CC on 20-MAR-2003 to correct DR field.)
 XX
 SQ Sequence 639 AA;
 Query Match 96.9%; Score 2178; DB 2; Length 639;
 Best Local Similarity 96.1%; Pred. No. 4.4e-147;
 Matches 417; Conservative 14; Mismatches 3; Indels 0; Gaps 0;
 QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTGRNDSMHEAFRGKITALVALGRTN 60
 DB 206 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTGRNDSMHEAFRGKITALVALGRTN 265
 QY 61 NANDTNGHGHVAGSVLGNSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSAQYS 120
 DB 266 NANDTNGHGHVAGSVLGNSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSAQYS 325
 QY 121 AGARHTNSGVAAGVNGAYTTDSRNDDYVRKNDMTILPAAGNEGNGGTTISAPGAKNAI 180
 DB 326 AGARHTNSGVAAGVNGAYTTDSRNDDYVRKNDMTILPAAGNERNGGTTISAPGAKNAI 385
 QY 181 TVGATENLRFSGSYADNINHVAFSSRGFTKDGRIKPDVWAPGTFILSARSSLPDSSF 240
 DB 386 TVGATENLRFSGSYADNINHVAFSSRGFTKDGRIKPDVWAPGTFILSARSSLPDSSF 445

241 WANHDSKYAYVGGTSMATPIVAGNVAOLREHFVKNRGITPKPILLKAAALAGADIGLGY 300
 DB 446 WANHDSKYAYVGGTSMATPIVAGNVAOLREHFVKNRGITPKPILLKAAALAGADIGLGY 505
 QY 301 PNGNQGWGRVTLDKSLNVAYVNESSLSLSTSQKATYSFTATAGKPLKISLVMSDAPASTTA 360
 DB 506 PNGNQGWGRVTLDKSLNVAYVNESSLSLSTSQKATYSFTATAGKPLKISLVMSDAPASTTA 565
 QY 361 SVTLVNDLNLVITAPNGTQYVGNDFSPYNDNDGNNVENVFNAPQSCTYTIIEVOAYN 420
 DB 566 SVTLVNDLNLVITAPNGTRYVGNDFSAFFDNNDGNNVENVFNAPQSCTYTIIEVOAYN 625
 QY 421 VPVGPQTFSLAIVN 434
 DB 626 VPVGPQNFSLAIVN 639
 RESULT 6
 AAY17087
 ID AAY17087 standard; protein; 639 AA.
 XX
 AC AAY17087;
 XX
 DT 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 DE An alkaline protease sequence from Bacillus species.
 XX
 KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent.
 XX
 OS Bacillus sp.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1. .639
 FT /note= "all residues indicated as Xaa are arbitrary amino
 FT acids"
 XX
 PN WO9918218-A1.
 XX
 PD 15-APR-1999.
 XX
 PF 07-OCT-1998; 98WO-JP004528.
 XX
 PR 07-OCT-1997; 97JP-00274570.
 XX
 PA (KAOS) KAO CORP.
 XX
 PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX
 DR WPI; 1999-287736/27.
 DR N-PSDB; AAX37277.
 XX
 PT Alkali protease from Bacillus used in washing powders.
 XX
 PS Claim 3; Page 47-50; 71pp; Japanese.
 XX
 CC The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease of the
 CC invention. (Updated on 20-MAR-2003 to correct DR field.)

```

XX SQ Sequence 639 AA;
Query Match 95.7%; Score 2150; DB 2; Length 639;
Best Local Similarity 96.1%; Pred. No. 4.4e-145;
Matches 417; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

Qy 1 NDVARGIVKADVAGSSVGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Db 206 NDVARGIVKADVAGSSVGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 265
Qy 61 NANTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
Db 266 NANTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 325
Qy 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNPGNGGTTISAPGTAKNAI 180
Db 326 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNPGNGGTTISAPGTAKNAI 385
Qy 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKDPVMAPTFILSARSSLAPDSSF 240
Db 386 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKDPVMAPTFILSARSSLAPDSSF 445
Qy 241 WANHDSKYAYMGCTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKALITAGADTGLGY 300
Db 446 WANHDSKYAYMGCTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKALITAGADTGLGY 505
Qy 301 PNGNQGWGRVTLDKSLNVAAYVNESSLSSTSQKATYFTATAGKPLKISLVWSDAPASTTA 360
Db 506 PNGNQGWGRVTLDKSLNVAAYVNESSLSSTSQKATYFTATAGKPLKISLVWSDAPASTTA 565
Qy 361 SVTLVNDLNLVITAPNGTQYVGNDFTPSYNDWDGRNNVENVFNAPOSQGTYYTIEVQAYN 420
Db 566 SVTLVNDLNLVITAPNGTQYVGNDFTPSYNDWDGRNNVENVFNAPOSQGTYYTIEVQAYN 625
Qy 421 VPVGPQTFSLAIVN 434
Db 626 VPVGPQTFSLAIVN 639

RESULT 7
AAV17088
ID AAV17088 standard; protein; 640 AA.
AC AAV17088;
DT 20-MAR-2003 (revised)
DT 21-JUL-1999 (first entry)
XX
DE An alkaline protease sequence from Bacillus species.
KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
KW washing composition; oxidising agent.
XX
OS Bacillus sp.
XX
FH Key Location/Qualifiers
FT Misc-difference 1..640
FT /note= "all residues indicated as Xaa are arbitrary amino
FT acids"
XX
XX WO9918218-A1.
PN
XX
PD 15-APR-1999.
XX
PF 07-OCT-1998; 98WO-JP004528.
XX
PR 07-OCT-1997; 97JP-00274570.
XX
PA (KAOS ) KAO CORP.
PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
PI Shikata S, Nomura M;

```

```

XX WPI: 1999-287736/27.
DR N-PSDB; AAX37278.
XX
PT Alkali protease from Bacillus used in washing powders.
XX
PS Claim 3; Page 50-53; 71pp; Japanese.
XX
CC The invention relates to alkaline proteases produced by strains of
CC Bacillus. The proteases ability to digest casein is not inhibited by
CC oleic acid and they have a high stability to oxidising agents. The
CC alkaline protease of the invention has the following properties: (a) it
CC is active over the pH range 4-13 and has at least 80% of its optimum
CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
CC its ability to digest casein is not inhibited by oleic acid; (e) it has
CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
CC used as enzymes in washing compositions for use in automatic dishwashers
CC and for washing clothes. The stability to oxidising agents allows the
CC enzyme to be an effective component of washing compositions including
CC bleaches. The present sequence represents an alkaline protease of the
CC invention. (Updated on 20-MAR-2003 to correct DR field.)
XX
SQ Sequence 640 AA;
Query Match 95.7%; Score 2150; DB 2; Length 640;
Best Local Similarity 96.1%; Pred. No. 4.5e-145;
Matches 417; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

Qy 1 NDVARGIVKADVAGSSVGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Db 207 NDVARGIVKADVAGSSVGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 266
Qy 61 NANTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
Db 267 NANTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 326
Qy 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNPGNGGTTISAPGTAKNAI 180
Db 327 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNPGNGGTTISAPGTAKNAI 386
Qy 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKDPVMAPTFILSARSSLAPDSSF 240
Db 387 TVGATENLRPSFGSYADNINHVAFSSRGPTKGRIPKDPVMAPTFILSARSSLAPDSSF 446
Qy 241 WANHDSKYAYMGCTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKALITAGADTGLGY 300
Db 447 WANHDSKYAYMGCTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKALITAGADTGLGY 506
Qy 301 PNGNQGWGRVTLDKSLNVAAYVNESSLSSTSQKATYFTATAGKPLKISLVWSDAPASTTA 360
Db 507 PNGNQGWGRVTLDKSLNVAAYVNESSLSSTSQKATYFTATAGKPLKISLVWSDAPASTTA 566
Qy 361 SVTLVNDLNLVITAPNGTQYVGNDFTPSYNDWDGRNNVENVFNAPOSQGTYYTIEVQAYN 420
Db 567 SVTLVNDLNLVITAPNGTQYVGNDFTPSYNDWDGRNNVENVFNAPOSQGTYYTIEVQAYN 626
Qy 421 VPVGPQTFSLAIVN 434
Db 627 VPVGPQTFSLAIVN 640

RESULT 8
AAM50085
ID AAM50085 standard; protein; 434 AA.
XX
XX AAM50085;
AC AAM50085;
XX
DT 12-AUG-2002 (first entry)
DE Bacillus sp alkaline protease protein A-1 fragment.
XX
XX Alkaline protease; detergent; laundry; bleaching; dishwasher.
KW

```

```

XX OS Bacillus sp.
XX PN EP1209233-A2.
XX PD 29-MAY-2002.
XX PF 22-NOV-2001; 2001EP-00127851.
XX PR 22-NOV-2000; 2000JP-00355166.
XX PR 12-APR-2001; 2001JP-00114048.
XX PA (KAOS ) KAO CORP.
XX PI Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
XX PI Okuda M, Saeki K;
XX DR WPI; 2002-437518/47.
XX PT New modified alkaline proteases useful in detergent compositions.
XX PS Claim 5; Page 18-19; 25pp; English.
XX CC This invention describes novel Bacillus sp. alkaline proteases useful in
XX CC detergent compositions, especially in laundry, bleaching or automatic
XX CC dishwasher detergents. The novel proteases have an increased detergency &
XX CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
XX CC sequence represents a fragment of the alkaline protease A-1 from Bacillus
XX CC sp NCIB12289 described in the method of the invention
XX SQ Sequence 434 AA;
Query Match 95.1%; Score 2138; DB 5; Length 434;
Best Local Similarity 93.3%; Pred. No. 1.9e-144;
Matches 405; Conservative 20; Mismatches 9; Indels 0; Gaps 0;
QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHAEPRGKITAIYALGRTN 60
DB 1 NDVARGIVKADVAQSSYGLYGQGVVAVADTGLDGRNDSMHAEPRGKITAIYALGRTN 60
QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 61 NANDPNGHGHVAGSVLNGSTNKGMAPQANLVFQSVMDNSGGLGGLPSNVSTLFSQAYS 120
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
DB 121 AGARIHTNSGAPVNGAYTTDSRNVDYVRKNDMAVLFAAGNEGPNGGTISAPGTAKNAI 180
QY 181 TVGATENLRPSFGSYADNINHVAFQSSRGPTKDGRIKPDVMAPTFILSARSSLAPDSF 240
DB 181 TVGATENLRPSFGSYADNINHVAFQSSRGPTKDGRIKPDVMAPTFILSARSSLAPDSF 240
QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
DB 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
QY 301 PNGNQGWGRVTLKSLNVAVYVNESSLSSTQKATYSFTATAGKPLKISLVWSDAPASTTA 360
DB 301 PNGNQGWGRVTLKSLNVAVFVNETSPLSTQKATYSFTAQAGKPLKISLVWSDAPGSTTA 360
QY 361 SVTLVNDLNLVITAPNGTQVYVNDFTSPYNDNWDGRNNVENVFVINAPOSCTYIEVQAYN 420
DB 361 SVTLVNDLNLVITAPNGTKYVYVNDFTAPYDNNWDGRNNVENVFVINAPOSCTYIEVQAYN 420
QY 421 VPVGQPTFSLAIVN 434
DB 421 VPQGPQAFSLAIVN 434
RESULT 9
AAW50086
ID AAW50086 standard; protein: 433 AA.
XX

```

```

AC AAW50086;
XX 12-AUG-2002 (first entry)
XX DE Bacillus sp alkaline protease protein A-2 fragment.
XX KW Alkaline protease; detergent; laundry; bleaching; dishwasher.
XX OS Bacillus sp.
XX PN EP1209233-A2.
XX PD 29-MAY-2002.
XX PF 22-NOV-2001; 2001EP-00127851.
XX PR 22-NOV-2000; 2000JP-00355166.
XX PR 12-APR-2001; 2001JP-00114048.
XX PA (KAOS ) KAO CORP.
XX PI Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
XX PI Okuda M, Saeki K;
XX DR WPI; 2002-437518/47.
XX PT New modified alkaline proteases useful in detergent compositions.
XX PS Claim 5; Page 20-21; 25pp; English.
XX CC This invention describes novel Bacillus sp. alkaline proteases useful in
XX CC detergent compositions, especially in laundry, bleaching or automatic
XX CC dishwasher detergents. The novel proteases have an increased detergency &
XX CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
XX CC sequence represents a fragment of the alkaline protease A-2 from Bacillus
XX CC sp NCIB12513 described in the method of the invention
XX SQ Sequence 433 AA;
Query Match 94.4%; Score 2120.5; DB 5; Length 433;
Best Local Similarity 93.3%; Pred. No. 3.4e-143;
Matches 405; Conservative 20; Mismatches 8; Indels 1; Gaps 1;
QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHAEPRGKITAIYALGRTN 60
DB 1 NDVARGIVKADVAQSNFGLYGQGIIVAVADTGLDGRNDSMHAEPRGKITAIYALGRTN 60
QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 61 NANDPNGHGHVAGSVLNG-ATNKGMAPQANLVFQSIMDSGGGLGGLPANLQTLFSQAYS 119
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
DB 120 AGARIHTNSGAPVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADNINHVAFQSSRGPTKDGRIKPDVMAPTFILSARSSLAPDSF 240
DB 180 TVGATENLRPSFGSYADNINHVAFQSSRGPTKDGRIKPDVMAPTFILSARSSLAPDSF 239
QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
DB 240 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADVGLGF 299
QY 301 PNGNQGWGRVTLKSLNVAVYVNESSLSSTQKATYSFTATAGKPLKISLVWSDAPASTTA 360
DB 300 PNGNQGWGRVTLKSLNVAVFVNETSPLSTQKATYSFTAQAGKPLKISLVWSDAPGSTTA 359
QY 361 SVTLVNDLNLVITAPNGTQVYVNDFTSPYNDNWDGRNNVENVFVINAPOSCTYIEVQAYN 420
DB 360 SLTLVNDLNLVITAPNGTKYVYVNDFTAPYDNNWDGRNNVENVFVINAPOSCTYIEVQAYN 419
QY 421 VPVGQPTFSLAIVN 434

```

Db	420	VPVSPQTFSLAIVH	433	
Db	269	NANDPNGHGTAVGSLVGN-ATNKGMAPQANLVFQSI	MDSGGLGGLPANLQTLFSQAYS	327
Db	121	AGARIHTNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTTISAPGTAKNAI		180
Db	328	AGARIHTNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTTISAPGTAKNAI		387
Db	181	TVGATENLRPSFGSYADNINHVAFQSSRGTPKGRIPKPDVMAFGTFFILSARSSILAPDSSF		240
Db	388	TVGATENLRPSFGSYADNINHVAFQSSRGTPKGRIPKPDVMAFGTFFILSARSSILAPDSSF		447
Db	241	WANHDSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAALIAAGADIGLGY		300
Db	448	WANHDSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAALIAAGADIGLGY		507
Db	301	PNGQGWGRVTLDKSLNVAYVNESSLSSTQKATYSFTATAGKPLKISLVNSDAPASTTA		360
Db	508	PNGQGWGRVTLDKSLNVAYVNESSLSSTQKATYSFTATAGKPLKISLVNSDAPASTTA		567
Db	361	SVTLVNDLNLVITAPNGTCVYVNDFTPTNDNWDGRNNVENVFAPQSGTYTIEVQAYN		420
Db	568	SVTLVNDLNLVITAPNGTCVYVNDFTPTNDNWDGRNNVENVFAPQSGTYTIEVQAYN		627
Db	421	VPVSPQTFSLAIVN	434	
Db	628	VPVSPQTFSLAIVH	641	
Db	420	VPVSPQTFSLAIVH	433	
Db	17-DEC-1998.			
Db	09-JUN-1998;	98WO-USO12005.		
Db	12-JUN-1997;	97US-00873479.		
Db	(NOVO)	NOVO NORDISK BIOTECH INC.		
Db	Sioma A,	Christianson L;		
Db	WPT;	1999-080908/07.		
Db	N-PSDB;	AAV82382.		
Db	Novel protease from Bacillus subtilis LC20 - useful in laundry and dishwashing detergents and for leather processing.			
Db	Claim 7;	Page 53-54; 77pp; English.		
Db	This is the amino acid sequence of a novel protease of Bacillus sp. JPI170 (NCIB 12513), as deduced from the nucleotide sequence of an isolated gene (see AAV82382). The entire protein, including the signal peptide and prepro region, has 77% identity to alkaline protease Y (see AAW89548) from Bacillus. The invention provides vectors, recombinant host cells and methods for the recombinant production of the protease. The protease is used in laundry and dishwashing detergents, for institutional and industrial cleaning, and for leather processing, as well as for debittering and enhancing the degree of hydrolysis of protein hydrolysates, for flavour development through hydrolysis of proteins, degradation of undesired peptides and in enzymatic synthesis of peptides. It has enhanced stability towards oxidation under alkaline conditions, e.g. towards bleaching agents of the peroxy type. The invention also provides mutant cells in which the protease activity is diminished. Such cells can be used for the production of heterologous recombinant proteins			
Db	Query Match	94.4%;	Score 2120.5;	DB 2; Length 641;
Db	Best Local Similarity	93.3%;	Pred. No. 5.7e-143;	
Db	Matches 405;	Conservative 20;	Mismatches 8;	Indels 1; Gaps 1;
Db	1 NDVARGIKADYAGSSVGLYCGQIVAVADTGTGNDSSMHEAFKGTALYALGRTN	60		
Db	209 NDVARGIKADYAGSSVGLYCGQIVAVADTGTGNDSSMHEAFKGTALYALGRTN	268		
Db	61 NANDTNGHGTAVGSLVGNSTNKGMAPQANLVFQSI	MDSGGLGGLPANLQTLFSQAYS	120	

CC detergent compositions, especially in laundry, bleaching or automatic
 CC dishwasher detergents. The novel proteases have an increased detergency &
 CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
 CC sequence represents a fragment of the alkaline protease SD-521 from
 CC Bacillus sp strain SD-521 (FERM BP-11162) described in the method of the
 CC invention

XX SQ Sequence 433 AA;

Query Match 88.7%; Score 1993.5; DB 5; Length 433;
 Best Local Similarity 87.6%; Pred. No. 4.1e-134;
 Matches 380; Conservative 29; Mismatches 24; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTGRNDSMHEAFRGKITALVALGRN 60
 DB 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDTGRNDSMHEAFRGKITALVALGRN 60
 QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
 DB 61 NANDPNGHGHVAGSVLGN-ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAWN 119
 QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
 DB 120 AGARIHTNSGAPVNGAYTANSRQVDEYVRNNDMTVLFAAGNEGPNNGTISAPGTAKNAI 179
 QY 181 TVGATENLRPSFGSYADNINHVAFSGRGTGRIKPDVMAPTGTFILSARSLAPDSF 240
 DB 180 TVGATENYRPSFGLADNPNHIAQFSRGATRGRIKPDVATGTFILSARSLAPDSF 239
 QY 241 WANHSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKALIAAGADIGLGY 300
 DB 240 WANYNSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKALIAAGADIGLGY 299
 QY 301 PNGNQGWGRVTLDKSLNVAVYNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
 DB 300 PSGDQGWGRVTLDKSLNVAVYNEATALTGQKATYSFQTAQKPLKISLVWTDAPGSTTA 359
 QY 421 VPVGPQTFSLAIVN 434
 DB 420 VPSPGQRFSLAIVH 433

RESULT 13

ID AAM50082 standard; protein; 433 AA.

XX AC AAM50082;

XX DT 12-AUG-2002 (first entry)

XX DE Bacillus sp D6-(FERM P1592) alkaline protease protein fragment.

XX KW Alkaline protease; detergent; laundry; bleaching; dishwasher.

XX OS Bacillus sp.

XX PN EP1209233-A2.

XX PD 29-MAY-2002.

XX PF 22-NOV-2001; 2001EP-00127851.

XX PR 22-NOV-2000; 2000JP-00355166.

XX PR 12-APR-2001; 2001JP-00114048.

XX XX (KACS) KAO CORP.

XX XX Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;

PI Okuda M, Saeki K;

XX WPI; 2002-437518/47.
 XX New modified alkaline proteases useful in detergent compositions.
 XX Claim 5; Page 13-15; 25pp; English.

CC This invention describes novel Bacillus sp. alkaline proteases useful in
 CC detergent compositions, especially in laundry, bleaching or automatic
 CC dishwasher detergents. The novel proteases have an increased detergency &
 CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
 CC sequence represents a fragment of the alkaline protease E-1 from Bacillus
 CC sp strain D6-(FERM-P1592) described in the method of the invention

XX SQ Sequence 433 AA;

Query Match 88.5%; Score 1989.5; DB 5; Length 433;

Best Local Similarity 87.3%; Pred. No. 7.8e-134;

Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTGRNDSMHEAFRGKITALVALGRN 60
 DB 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDTGRNDSMHEAFRGKITALVALGRN 60
 QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
 DB 61 NANDPNGHGHVAGSVLGN-ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAWN 119
 QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
 DB 120 AGARIHTNSGAPVNGAYTANSRQVDEYVRNNDMTVLFAAGNEGPNNGTISAPGTAKNAI 179
 QY 181 TVGATENLRPSFGSYADNINHVAFSGRGTGRIKPDVMAPTGTFILSARSLAPDSF 240
 DB 180 TVGATENYRPSFGLADNPNHIAQFSRGATRGRIKPDVATGTFILSARSLAPDSF 239
 QY 241 WANHSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKALIAAGADIGLGY 300
 DB 240 WANYNSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKALIAAGADIGLGY 299
 QY 301 PNGNQGWGRVTLDKSLNVAVYNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
 DB 300 PSGDQGWGRVTLDKSLNVAVYNEATALTGQKATYSFQTAQKPLKISLVWTDAPGSTTA 359
 QY 361 SVTLVNDLNLVITAPNGQYVGNDFTSYNDNWDGRNENVFINAPQSGTYTIEVOAYN 420
 DB 360 SYTLVNDLNLVITAPNGQYVGNDFSYPDNNDGRNENVFINAPQSGTYTIEVOAYN 419
 QY 421 VPVGPQTFSLAIVN 434
 DB 420 VPSPGQRFSLAIVH 433

RESULT 14

AAM89548

ID AAM89548 standard; protein; 636 AA.

XX AC AAM89548;

XX DT 12-APR-1999 (first entry)

XX DE Bacillus sp. alkaline protease Y.

XX KW Alkaline protease Y; detergent; surfactant; leather processing;

XX KW debittering; flavour.

XX OS Bacillus sp.

XX PN WO9856927-A2.

XX PD 17-DEC-1998.

XX PF 09-JUN-1998; 98WO-US012005.

XX 12-JUN-1997; 97US-00873479.
 XX (NOVO) NOVO NORDISK BIOTECH INC.
 PA Sloma A, Christianson L;
 XX WPI; 1999-080908/07.
 XX Novel protease from *Bacillus subtilis* LC20 - useful in laundry and
 PT dishwashing detergents and for leather processing.
 PT Claim 3; Page 55-56; 77pp; English.
 XX This is the amino acid sequence of a *Bacillus* sp. alkaline protease Y
 CC that is said to have good alkali and surfactant resistance and improved
 CC detergency. It shows 77% identity to a newly isolated protease (see
 CC AA089547) of *Bacillus* sp. JP170 (NCIB 12513). The invention provides
 CC vectors, recombinant host cells and methods for the recombinant
 CC production of such proteases. The protease are used in laundry and
 CC dishwashing detergents, for institutional and industrial cleaning, and
 CC for leather processing, as well as for debittering and enhancing the
 CC degree of hydrolysis of protein hydrolysates, for flavour development
 CC through hydrolysis of proteins, degradation of undesired peptides and in
 CC enzymatic synthesis of peptides. They have enhanced stability towards
 CC oxidation under alkaline conditions, e.g. towards bleaching agents of the
 CC peroxy type. The invention also provides mutant cells in which the
 CC protease activity is diminished. Such cells can be used for the
 CC production of heterologous recombinant proteins
 XX Sequence 636 AA;
 XX
 Query Match 88.3%; Score 1984.5; DB 2; Length 636;
 Best Local Similarity 87.1%; Pred. No. 2.9e-133;
 Matches 378; Conservative 31; Mismatches 24; Indels 1; Gaps 1;
 QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITIYALGRIN 60
 DB 204 NDVARGIVKADVAQNNYGLYGQQLVAVADTGLDTRNDSSMHEAFRGKITIYALGRIN 263
 QY 61 NANDTNGHGHVAGSVLNGSTKNGAPQANLVFQSIMDSGGGLGGLPSNLTLFSQAYS 120
 DB 264 NASDPNGHGHVAGSVLGN-ALNKGMAPQANLVFQSIMDSGGGLGGLPSNLTLFSQAWN 322
 QY 121 AGARIHTNSGAAVNGAYTTDSNRVDYVRKNDMTILFAAGNEGPNCGTISAPGTAKNAI 180
 DB 323 AGARIHTNSGAPVNGAYTANSROVDYVRNNDMTVLFAAGNEGPNSGTISAPGTAKNAI 382
 QY 181 TVGATENLRPSFGSYADNINEVAQFSRGPDKGRIPKDVMAQPTPILSARSLAPDSSF 240
 DB 383 TVGATENYRPSFGSIADNPNHIAQFSRGATRDGRIPKDVTAQPTPILSARSLAPDSSF 442
 QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGADIGLY 300
 DB 443 WANYNSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGADIGLY 502
 QY 301 PNGNQGWGRVTLDKSLNVAAYVNESSLSQKATYSTATAGKPKLSLVWSDAPASTTA 360
 DB 503 PSQDQGWGRVTLDKSLNVAAYVNEATALATGQKATYSFQAQKPKLSLVWTDAPGSTTA 562
 QY 361 SVTLVNDLNLVITAPNGTQVYVGNDFSPYNDNDGRNNVNFVFNAPQSGTYYIEVOAYN 420
 DB 563 SYTLVNDLNLVITAPNGQKYVGNDFSAFYDNNWDGRNNVNFVFNAPQSGTYYIEVOAYN 622
 QY 421 VPVGPQTFFSLAIVN 434
 DB 623 VPSPGFQFSLAIVH 636
 RESULT 15
 AAM50083
 ID AAM50083 standard; protein; 433 AA.
 XX

AC AAM50083;
 XX 12-AUG-2002 (first entry)
 DT Bacillus sp Y-(FERM BP-1029) alkaline protease protein fragment.
 DE Alkaline protease; detergent; laundry; bleaching; dishwasher.
 XX Alkaline protease; detergent; laundry; bleaching; dishwasher.
 XX Bacillus sp.
 OS EPI209233-A2.
 PN 29-MAY-2002.
 PD 22-NOV-2001; 2001EP-00127851.
 PF 22-NOV-2000; 2000JP-00355166.
 PR 12-APR-2001; 2001JP-00114048.
 XX (KAOS) KAO CORP.
 PA Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
 PI Okuda M, Saeki K;
 FI WPI; 2002-437518/47.
 DR New modified alkaline proteases useful in detergent compositions.
 XX Claim 5; Page 15-16; 25pp; English.
 PS This invention describes novel *Bacillus* sp. alkaline proteases useful in
 XX detergent compositions, especially in laundry, bleaching or automatic
 CC dishwasher detergents. The novel proteases have an increased detergency &
 CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
 CC sequence represents a fragment of the alkaline protease Ya from *Bacillus*
 CC sp strain Y-(FERM BP-1029) described in the method of the invention
 XX Sequence 433 AA;
 XX
 Query Match 88.2%; Score 1982.5; DB 5; Length 433;
 Best Local Similarity 87.1%; Pred. No. 2.5e-133;
 Matches 378; Conservative 30; Mismatches 25; Indels 1; Gaps 1;
 QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITIYALGRIN 60
 DB 1 NDVARGIVKADVAQNNYGLYGQQLVAVADTGLDTRNDSSMHEAFRGKITIYALGRIN 60
 QY 61 NANDTNGHGHVAGSVLNGSTKNGAPQANLVFQSIMDSGGGLGGLPSNLTLFSQAYS 120
 DB 61 NASDPNGHGHVAGSVLGN-ALNKGMAPQANLVFQSIMDSGGGLGGLPSNLTLFSQAWN 119
 QY 121 AGARIHTNSGAAVNGAYTTDSNRVDYVRKNDMTILFAAGNEGPNCGTISAPGTAKNAI 180
 DB 120 AGARIHTNSGAPVNGAYTANSROVDYVRNNDMTVLFAAGNEGPNSGTISAPGTAKNAI 179
 QY 181 TVGATENLRPSFGSYADNINEVAQFSRGPDKGRIPKDVMAQPTPILSARSLAPDSSF 240
 DB 180 TVGATENYRPSFGSIADNPNHIAQFSRGATRDGRIPKDVTAQPTPILSARSLAPDSSF 239
 QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGADIGLY 300
 DB 240 WANYNSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGADIGLY 299
 QY 301 PNGNQGWGRVTLDKSLNVAAYVNESSLSQKATYSTATAGKPKLSLVWSDAPASTTA 360
 DB 300 PSQDQGWGRVTLDKSLNVAAYVNEATALATGQKATYSFQAQKPKLSLVWTDAPGSTTA 359
 QY 361 SVTLVNDLNLVITAPNGTQVYVGNDFSPYNDNDGRNNVNFVFNAPQSGTYYIEVOAYN 420
 DB 360 SYTLVNDLNLVITAPNGQKYVGNDFSAFYDNNWDGRNNVNFVFNAPQSGTYYIEVOAYN 419
 QY 421 VPVGPQTFFSLAIVN 434
 XX

Db 420 VPSGPQRFSLAIVH 433

Search completed: March 10, 2004, 14:53:09
Job time : 60 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:52:04 ; Search time 23 Seconds
(without alignments)
974.160 Million cell updates/sec

Title: US-09-985-689A-1-COPY
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAYNPVGPOTFSLAIVN 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/2/iaa/PCPUS COMB.pcp.*
6: /cgn2_6/ptodata/2/iaa/backfile1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2242	99.8	640	4	US-09-509-814A-6
2	2237	99.6	640	4	US-09-509-814A-8
3	2178	96.9	639	4	US-09-509-814A-4
4	2150	95.7	639	4	US-09-509-814A-1
5	2150	95.7	640	4	US-09-509-814A-2
6	2120.5	94.4	641	2	US-08-873-479-42
7	1981.5	88.2	433	4	US-09-104-623A-4
8	1981.5	88.2	433	4	US-09-019-532-4
9	1981.5	88.2	433	4	US-08-338-746-4
10	1981.5	88.2	435	2	US-08-873-479-43
11	1581.5	70.4	345	4	US-08-512-251A-10
12	1581.5	70.4	345	4	US-09-515-150A-10
13	1581.5	70.4	345	4	US-09-196-281-13
14	447.5	19.9	659	3	US-08-894-818B-1
15	447.5	19.9	659	3	US-09-445-472-12
16	409	18.2	412	4	US-09-445-472-1
17	409	18.2	522	3	US-08-894-818B-3
18	409	18.2	522	4	US-09-445-472-4
19	409	18.2	654	3	US-08-894-818B-35
20	409	18.2	654	4	US-09-445-472-16
21	396	17.6	659	3	US-08-894-818B-5
22	348	15.5	520	3	US-09-000-016-7
23	348	15.5	520	4	US-09-514-340-7
24	348	15.5	734	3	US-09-000-016-4
25	348	15.5	734	4	US-09-514-340-4
26	348	15.5	823	3	US-09-000-016-2
27	348	15.5	823	4	US-09-514-340-2

28	305.5	13.6	903	1	US-08-750-532-1	Sequence 1, Appli
29	305.5	13.6	1398	1	US-08-750-532-9	Sequence 9, Appli
30	305.5	13.6	1398	3	US-08-894-818B-8	Sequence 8, Appli
31	305.5	13.6	1398	4	US-09-445-472-6	Sequence 6, Appli
32	295.5	13.2	237	1	US-08-750-532-18	Sequence 18, Appli
33	283.5	12.6	418	2	US-08-873-479-44	Sequence 44, Appli
34	276	12.3	418	4	US-09-966-921A-2	Sequence 2, Appli
35	256.5	11.4	397	4	US-09-328-352-7533	Sequence 7533, Ap
36	254.5	11.3	275	1	US-08-431-387-1	Sequence 1, Appli
37	251	11.2	280	1	US-08-434-255-8	Sequence 8, Appli
38	251	11.2	280	1	US-08-459-967-8	Sequence 8, Appli
39	251	11.2	280	1	US-08-460-327-8	Sequence 8, Appli
40	251	11.2	280	1	US-08-459-871-8	Sequence 8, Appli
41	251	11.2	280	3	US-09-024-532-2	Sequence 2, Appli
42	251	11.2	280	4	US-09-104-623A-2	Sequence 2, Appli
43	251	11.2	280	4	US-09-019-532-2	Sequence 2, Appli
44	251	11.2	280	4	US-09-417-359A-2	Sequence 2, Appli
45	251	11.2	280	4	US-09-705-185-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-509-814A-6
; Sequence 6, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAKKI, KATSUHIKA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, MASATUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-OPCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 6
; LENGTH: 640
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-509-814A-6

Query Match	99.8%	Score 2242;	DB 4;	Length 640;
Best Local Similarity	99.8%	Pred. No. 7.4e-172;		
Matches 433;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	NDVARGIVKADVAQSSYGLYGGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN	60	
Db	207	NDVARGIVKADVAQSSYGLYGGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN	266	
QY	61	NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS	120	
Db	267	NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS	326	
QY	121	AGARHTNSWGAANGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI	180	
Db	327	AGARHTNSWGAANGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI	386	
QY	181	TVGATENLRPFGSYADNINHVAQFSRRGPTKDGRIKPDVMAFGTIFLSARSLAPDSF	240	
Db	387	TVGATENLRPFGSYADNINHVAQFSRRGPTKDGRIKPDVMAFGTIFLSARSLAPDSF	446	
QY	241	WANHDSKIAYMGGTSMTPIVAGNVAQLREHFVKNRGITPKPSLLKALIAAGAADIGLY	300	

Db 447 WANHDSKYAYMGSTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAGAADIGLY 506
Qy 301 PNGNQGWGRVTLDKSLNVAAYNNESSLSSTQKATYSFTATAGKPKLSILVWSDAPASTTA 360
Db 507 PNGNQGWGRVTLDKSLNVAAYNNESSLSSTQKATYSFTATAGKPKLSILVWSDAPASTTA 566
Qy 361 SVTLVNDLNLVITAPNGTQYVGNDFTPYNDNWDGRNNVNFVFINAPQSGTYTIEVQAYN 420
Db 567 SVTLVNDLNLVITAPNGTQYVGNDFTPYNDNWDGRNNVNFVFINAPQSGTYTIEVQAYN 626
Qy 421 VPVGPQTFSLAIVN 434
Db 627 VPVGPQTFSLAIVN 640

RESULT 2

US-09-509-814A-8
; Sequence 8, Application US/09509814A

; Patent No. 6376227

; GENERAL INFORMATION:

; APPLICANT: TAKAIWA, MIKIO

; APPLICANT: OKUDA, MITSUYOSHI

; APPLICANT: SAEKI, KATSUHIISA

; APPLICANT: KUBOTA, HIROMI

; APPLICANT: HITOMI, JUN

; APPLICANT: KAGEYAMA, YASUSHI

; APPLICANT: SHIKATA, SHITSUM

; APPLICANT: NOMURA, MASAFUMI

; TITLE OF INVENTION: ALKALINE PROTEASE

; FILE REFERENCE: 0327-0832-OPCT

; CURRENT APPLICATION NUMBER: US/09/509,814A

; CURRENT FILING DATE: 2000-04-06

; PRIOR APPLICATION NUMBER: PCT/JP98/04528

; PRIOR FILING DATE: 1998-10-07

; PRIOR APPLICATION NUMBER: JP 9-274570

; PRIOR FILING DATE: 1997-06-08

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Patent in version 3.0

; SEQ ID NO 8

; LENGTH: 640

; TYPE: PRT

; ORGANISM: Bacillus sp.

US-09-509-814A-8

Query Match 99.6%; Score 2237; DB 4; Length 640;

Best Local Similarity 99.5%; Pred. No. 1.9e-171; Mismatches 1; Indels 0; Gaps 0;

Matches 432; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHAEFRGKITALYALGRTN 60
Db 207 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHAEFRGKITALYALGRTN 266
Qy 61 NANDTNGHGHVAGSVLNGSTNKGWAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 267 NANDTNGHGHVAGSVLNGSTNKGWAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 326
Qy 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 327 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 386
Qy 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTILSARSLAPDSF 240
Db 387 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTILSARSLAPDSF 446
Qy 241 WANHDSKYAYMGSTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAGAADIGLY 300
Db 447 WANHDSKYAYMGSTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAGAADIGLY 506
Qy 301 PNGNQGWGRVTLDKSLNVAAYNNESSLSSTQKATYSFTATAGKPKLSILVWSDAPASTTA 360
Db 507 PNGNQGWGRVTLDKSLNVAAYNNESSLSSTQKATYSFTATAGKPKLSILVWSDAPASTTA 566

Qy 361 SVTLVNDLNLVITAPNGTQYVGNDFTPYNDNWDGRNNVNFVFINAPQSGTYTIEVQAYN 420
Db 567 SVTLVNDLNLVITAPNGTQYVGNDFTPYNDNWDGRNNVNFVFINAPQSGTYTIEVQAYN 626
Qy 421 VPVGPQTFSLAIVN 434
Db 627 VPVGPQTFSLAIVN 640

RESULT 3

US-09-509-814A-4

; Sequence 4, Application US/09509814A

; Patent No. 6376227

; GENERAL INFORMATION:

; APPLICANT: TAKAIWA, MIKIO

; APPLICANT: OKUDA, MITSUYOSHI

; APPLICANT: SAEKI, KATSUHIISA

; APPLICANT: KUBOTA, HIROMI

; APPLICANT: HITOMI, JUN

; APPLICANT: KAGEYAMA, YASUSHI

; APPLICANT: SHIKATA, SHITSUM

; APPLICANT: NOMURA, MASAFUMI

; TITLE OF INVENTION: ALKALINE PROTEASE

; FILE REFERENCE: 0327-0832-OPCT

; CURRENT APPLICATION NUMBER: US/09/509,814A

; CURRENT FILING DATE: 2000-04-06

; PRIOR APPLICATION NUMBER: PCT/JP98/04528

; PRIOR FILING DATE: 1998-10-07

; PRIOR APPLICATION NUMBER: JP 9-274570

; PRIOR FILING DATE: 1997-06-08

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Patent in version 3.0

; SEQ ID NO 4

; LENGTH: 639

; TYPE: PRT

; ORGANISM: Bacillus sp.

US-09-509-814A-4

Query Match 96.9%; Score 2178; DB 4; Length 639;

Best Local Similarity 96.1%; Pred. No. 1e-166; Mismatches 3; Indels 0; Gaps 0;

Matches 417; Conservative 14; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHAEFRGKITALYALGRTN 60
Db 206 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHAEFRGKITALYALGRTN 265
Qy 61 NANDTNGHGHVAGSVLNGSTNKGWAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 266 NANDTNGHGHVAGSVLNGSTNKGWAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 325
Qy 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 326 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 385
Qy 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTILSARSLAPDSF 240
Db 386 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTILSARSLAPDSF 445
Qy 241 WANHDSKYAYMGSTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAGAADIGLY 300
Db 446 WANHDSKYAYMGSTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAGAADIGLY 505
Qy 301 PNGNQGWGRVTLDKSLNVAAYNNESSLSSTQKATYSFTATAGKPKLSILVWSDAPASTTA 360
Db 506 PNGNQGWGRVTLDKSLNVAAYNNESSLSSTQKATYSFTATAGKPKLSILVWSDAPASTTA 565
Qy 361 SVTLVNDLNLVITAPNGTQYVGNDFTPYNDNWDGRNNVNFVFINAPQSGTYTIEVQAYN 420
Db 566 SVTLVNDLNLVITAPNGTQYVGNDFTPYNDNWDGRNNVNFVFINAPQSGTYTIEVQAYN 625
Qy 421 VPVGPQTFSLAIVN 434
Db 626 VPVGPQTFSLAIVN 639